

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:43:42 ; Search time 37.5566 Seconds
(without alignments)
1482.081 Million cell updates/sec

Title: US-09-980-403-2_COPY_159_355

Perfect score: 1022

Sequence: 1 CQTYMDIVIVLDGNSIYPW.....AALKDIVDALGDRIFSLGEGT 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1022	100.0	707	4 AAU19663	AAU19663 Human nov
2	1022	100.0	707	5 ABP47883	ABP47883 Human pol
3	1022	100.0	707	7 ADC10845	ADC10845 Human ext
4	1022	100.0	1034	3 AAB25590	AAB25590 Protein e
5	1022	100.0	1034	6 ADA27062	ADA27062 Human nov
6	1022	100.0	1034	8 ADE86592	ADE86592 Novel hum
7	1022	100.0	1120	6 ABR58365	ABR58365 Human NOV
8	1022	100.0	1188	4 AAB30929	AAB30929 Amino aci
9	1022	100.0	1188	4 AAU14467	AAU14467 Human nov
10	1022	100.0	1188	4 AAU14231	AAU14231 Human nov
11	1022	100.0	1188	4 AAB50085	AAB50085 Human A25
12	1022	100.0	1188	5 AAU10551	AAU10551 Human A25
13	1022	100.0	1188	7 ADE09956	ADE09956 Novel pro
14	1022	100.0	1189	3 AAB25582	AAB25582 ITGAL1 pr
15	1022	100.0	1189	4 AAB25582	ABG12949 Novel hum
16	1022	100.0	1189	6 ABR58364	ABR58364 Human NOV
17	1022	100.0	1189	6 ADA27054	ADA27054 Human nov
18	1022	100.0	1189	7 ADE63570	ADE63570 Human pro
19	1022	100.0	1189	8 ADE86584	ADE86584 Novel hum
20	1010	98.8	1188	4 AAB50087	AAB50087 Marine A2
21	1010	98.8	1188	5 AAU10552	AAU10552 Marine A2
22	1002	98.0	193	5 AAU76854	AAU76854 Human int
23	997	97.6	193	5 AAU76863	AAU76863 Human int
24	979	95.8	360	7 ADE08585	ADE08585 Novel pro
25	599	58.6	1151	8 ADE86652	ADE86652 Human int

ALIGNMENTS

RESULT 1

AAU19663

ID AAU19663 standard; protein; 707 AA.

XX

AC AAU19663;

XX

DT 04-DEC-2001 (first entry)

XX

DE Human novel extracellular matrix protein, Seq ID No 313.

XX

KW Human; secreted extracellular matrix protein; immunomodulatory; Anti-HIV;
KW antianemic; antirheumatic; antisclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW antiallergics; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancer; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.

OS Homo sapiens.

XX

PN WO200155368-A1.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US001348.

XX

PR 31-JAN-2000; 2000US-0179065P.

PR

PR 04-FEB-2000; 2000US-0180628P.

PR

PR 24-FEB-2000; 2000US-0184664P.

PR

PR 02-MAR-2000; 2000US-0186350P.

PR

PR 16-MAR-2000; 2000US-0189874P.

PR

PR 17-MAR-2000; 2000US-0190076P.

PR

PR 18-APR-2000; 2000US-0198123P.

PR

PR 19-MAY-2000; 2000US-0205515P.

PR

PR 07-JUN-2000; 2000US-0209467P.

PR

PR 28-JUN-2000; 2000US-0214886P.

PR

PR 30-JUL-2000; 2000US-0215135P.

PR

PR 07-JUL-2000; 2000US-0216647P.

PR

PR 07-JUL-2000; 2000US-0216800P.

PR

PR 11-JUL-2000; 2000US-0217487P.

PR

PR 11-JUL-2000; 2000US-0217496P.

PR

PR 14-JUL-2000; 2000US-0218290P.

PR

PR 26-JUL-2000; 2000US-0220963P.

PR

PR 26-JUL-2000; 2000US-0220964P.

PR

PR 14-AUG-2000; 2000US-0224518P.

PR

CC virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis),
 CC cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of
 CC the breast or liver, Sezary syndrome and Gaucher's disease), neurological
 CC diseases (e.g. Alzheimer's disease, Parkinson's disease), cardio-
 CC /cerebrovascular disorders (e.g. cardiac arrest, tachycardia and angina),
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC (e.g. corneal infections). Other uses include wound healing, maintenance
 CC of organs before transplantation, support of cell culture of primary

Query Match 100.0%; Score 1022; DB 4; Length 707;
 Best Local Similarity 100.0%; Pred. No. 8.3e-99;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 COTYMDIVIVLQGSNIYPWVEVQHFLLNKKFYIGPQIQGVVQYQGVGVVHFLND 60
 Db 175 COTYMDIVIVLQGSNIYPWVEVQHFLLNKKFYIGPQIQGVVQYQGVGVVHFLND 234

QY 61 YRSVKDVEAAHIEQFGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGHSDSP 120
 Db 235 YRSVKDVEAAHIEQFGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGHSDSP 294

QY 121 DLEKVIQOSERDNVRYAVAVILGYNNRGINPETFLNEIKYIASDDDDKHFFNVTDAAAL 180
 Db 295 DLEKVIQOSERDNVRYAVAVILGYNNRGINPETFLNEIKYIASDDDDKHFFNVTDAAAL 354

QY 181 KDIVDALGDRIFSLEGT 197
 Db 355 KDIVDALGDRIFSLEGT 371

RESULT 2
 ABP47883
 ID ABP47883 standard; protein; 707 AA.
 AC ABP47883;
 XX
 DT 23-AUG-2002 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 313.
 XX
 KW Human; neutropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
 KW antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatoprotective; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
 XX
 OS Homo sapiens.
 XX
 PN US2002042386-A1.
 XX
 PD 11-APR-2002.
 XX
 PF 17-JAN-2001; 2001US-00764870.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0225868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0237040P.
 PR 20-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 XX WPI; 2002-470713/50.
 DR N-PSDB; ABQ66558.
 XX
 PT New nucleic acid encoding human proteins, useful for diagnosis, treatment
 PT and prevention of e.g. osteoporosis, also related polypeptides and
 PT antibodies.
 XX
 PS Claim 11; SEQ ID NO 313; 235pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABQ66521-ABQ66785) and proteins
 CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html?docID=999909764870
 XX
 SQ Sequence 707 AA;

Query Match 100.0%; Score 1022; DB 5; Length 707;
 Best Local Similarity 100.0%; Pred. No. 8.3e-99;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQYMDIVLVLDGNSIYFPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVHFEHLND 60
Db |||||
175 CQYMDIVLVLDGNSIYFPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVHFEHLND 234
QY 61 YRSVKDVVEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 120
Db |||||
235 YRSVKDVVEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 294
QY 121 DLEKVIQQRDNVTRYAVAVLGYNRRGINPETFNEIKYIASDDDDKHFFNVTDAAAL 190
Db |||||
295 DLEKVIQQRDNVTRYAVAVLGYNRRGINPETFNEIKYIASDDDDKHFFNVTDAAAL 354
QY 181 KDIVDALGDRIFSLG 197
Db |||||
355 KDIVDALGDRIFSLG 371
RESULT 3
ADCL0845
ID ADC10845 standard; protein; 707 AA.
XX
AC ADC10845;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human extracellular matrix protein from gene 38.
XX
KW Extracellular matrix protein; cytostatic; antibacterial; virucide;
KW neuroprotective; gynaecological; gastrointestinal-Gen; cardiant;
KW cardiovascular-Gen; nephrotropic; antiinflammatory; muscular-Gen;
KW respiratory-Gen; immunosuppressive; cerebroprotective; vasotropic;
KW nootropic; antiallergic; cancer; bacterial infection; viral infection;
KW neural disorder; immune system disorder; blood disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW inflammatory disorder; proliferative disorder; Human.
XX
OS Homo sapiens.
XX
PN US2003059875-A1.
XX
PD 27-MAR-2003.
XX
PF 19-APR-2002; 2002US-00125540.
XX
PX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0251989P.
 PR 06-DEC-2000; 2000US-0256719P.
 PR 08-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251888P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PR 17-JAN-2001; 2001US-00764870.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-743765/70.
 XX N-PSDB; ADC10580.
 DR New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating, and/or preventing disorders, such as cancer, infections,
 PT cardiovascular and inflammatory diseases.
 XX Claim 11; SEQ ID NO 313; 235pp; English.
 PS The invention relates to an isolated nucleic acid molecule (cDNA)
 CC encoding a human extracellular matrix protein, representing one of 161
 CC novel genes. Also included are recombinant vectors, host cells
 CC (expressing the protein), the extracellular matrix proteins (including
 CC their fragments, epitopes and homologues), an isolated antibody that
 CC binds specifically to the protein, diagnosing a pathological condition or
 CC susceptibility to a pathological condition (comprising determining the
 CC presence or absence of a mutation in the nucleic acid and diagnosing a
 CC condition based on the presence or absence of the mutation), diagnosing a
 CC pathological condition or susceptibility to a pathological condition
 CC (comprising determining the presence or amount of expression of the
 CC protein in a biological sample and diagnosing a condition based on the
 CC presence or amount of expression of the protein), preventing, treating or
 CC ameliorating a medical condition by administering the nucleic acid or
 CC protein to a mammalian subject, identifying a binding partner to the
 CC protein, the gene corresponding to the cDNA sequence, and identifying an
 CC activity in a biological assay (comprising expressing the nucleic acid in
 CC a cell, isolating the supernatant, detecting an activity in a biological
 CC assay and identifying the protein in the supernatant having the
 CC activity). The nucleic acids and proteins display the following
 CC activities: Cytostatic, antibacterial, Virucide, Neuroprotective,
 CC Gynecological, Gastrointestinal-Gen, Cardiant, Cardiovascular-Gen,
 CC Nephrotropic, Antiinflammatory, Muscular-Gen, Respiratory-Gen,

Query Match 100.0%; Score 1022; DB 7; Length 707;
 Best Local Similarity 100.0%; Pred. No. 8.3e-99;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CQTYMDIVIVLDGNSIYPWVEVOHELINILKKFYIGPGQIQGVGVQYGVWVHEFHND 60
 |||||
 Db 175 CQTYMDIVIVLDGNSIYPWVEVOHELINILKKFYIGPGQIQGVGVQYGVWVHEFHND 234
 |||||
 QY 61 YRSVKDVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKKGAKKVMIVITDGESHDP 120
 |||||
 Db 235 YRSVKDVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKKGAKKVMIVITDGESHDP 294
 |||||
 QY 121 DLEKVIQOQSERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDDPKHFNVTDEAAL 180
 |||||
 Db 295 DLEKVIQOQSERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDDPKHFNVTDEAAL 354
 |||||
 QY 181 KDIVDALGDRIFSLEGT 197
 |||||
 Db 355 KDIVDALGDRIFSLEGT 371
 |||||
 RESULT 4
 AAB25590
 ID AAB25590 standard; protein; 1034 AA.
 XX
 AC AAB25590;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Protein encoded by human secreted protein gene #7 clone HOHBY69.
 XX
 KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
 KW anticancer; vulnary; antiviral; antibacterial; antifungal;
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KW Crohn's disease; nephritis; hyperproliferative disorder;
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
 XX
 OS Homo sapiens.
 XX
 PN WO2000029435-A1.
 XX
 PD 25-MAY-2000.
 XX
 PF 27-OCT-1999; 99WO-US025031.
 XX
 PR 28-OCT-1998; 98US-0105971P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
 XX Greene JM;
 XX WPI; 2000-387742/33.
 XX
 PT Isolated nucleic acid molecules encoding human secreted proteins are used
 PT for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 XX wounds, and infectious diseases.
 PS Claim 1; Page 678-682; 803pp; English.
 CC The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given in
 CC AAA80606-8A0623 encode the 12 secreted protein sequences given in
 CC AAB25576-B25593. The human secreted proteins have various activities
 CC dependent on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant; anti-inflammatory;
 CC antiarthritic; antirheumatic; dermatological; antiproliferative;
 CC antiarteriosclerotic; anticancer; vulnary; antiviral; antibacterial;

CC and antifungal activity. The proteins, polypeptides, agonists and
CC antagonists may be used to treat prevent and/or diagnose various disease,
CC disorders and conditions examples of which include: immune disorders e.g.
CC Addison's disease, rheumatoid arthritis, dermatitis, and multiple
CC sclerosis; inflammatory disorders e.g. inflammatory bowel disease,
CC Crohn's disease and nephritis; hyperproliferative disorders such as
CC paraproteinemias and purpura; cardiovascular disorders e.g. coronary
CC arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The
CC proteins and polynucleotide sequences may also be used in wound healing
CC and the treatment of infectious diseases. The human secreted protein gene
CC #7 and protein sequences are represented in sequences AAA80612 and
CC AAB2582. Secreted protein gene #7 is located at position chromosome 15
CC q22.3-23. Sequences AAA80652-A80661 represent genes which are related to
CC the secreted protein gene#7
XX
SQ Sequence 1034 AA;

Query Match 100.0%; Score 1022; DB 3; Length 1034;
Best Local Similarity 100.0%; Pred. No. 1.4e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQTYMDIVILVLDGNSIYPWVEVQHFLINILKFKYIGPGQIQVGWYQGVGVVHFEHLND 60
DB 159 CQTYMDIVILVLDGNSIYPWVEVQHFLINILKFKYIGPGQIQVGWYQGVGVVHFEHLND 218

QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGHSDSP 120
DB 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGHSDSP 278

QY 121 DLEKVIQOSERDNTVRYAVAVLGYNNRRGINPEFLNEIKYIASDPDDKHFNVTDAAAL 180
DB 279 DLEKVIQOSERDNTVRYAVAVLGYNNRRGINPEFLNEIKYIASDPDDKHFNVTDAAAL 338

QY 181 KDIVDALGDRIFSLEGT 197
DB 339 KDIVDALGDRIFSLEGT 355

RESULT 5
ADA27062
ID ADA27062 standard; protein; 1034 AA.
XX
AC ADA27062;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human novel secreted protein from cdna HOHBY69 #2.
XX
KW cytostatic; antiinflammatory; immunomodulator; neuroprotective;
KW hemostatic; gene therapy; cancer; inflammation; immune disorder;
KW neurological disorder; blood clotting disorder; food additive;
KW preservative; human; secreted protein.
XX
OS Homo sapiens.
XX
PN US2003055231-A1.
XX
PD 20-MAR-2003.
XX
PF 29-OCT-2001; 2001US-00984130.
XX
PR 28-OCT-1998; 98US-0105971P.
PR 27-OCT-1999; 99WO-0502503A.
PR 19-APR-2000; 2000US-0198407P.
PR 30-OCT-2000; 2000US-0243792P.
PR 18-APR-2001; 2001US-00836353.
XX
XX (NIJ/) NI J.
PA (YOUNG/) YOUNG P E.
PA (KERN/) KENNY J J.
PA (OLSEN/) OLSEN H S.
PA (MOOR/) MOORE P A.
PA (WEI/) WEI Y.

PA (GREE/) GREENE J M.
PA (RUBE/) RUBEN S M.
PA (LIUD/) LIU D.
PA (CROC/) CROCKER P R.
XX
PI Ni J, Young PE, Kenny JJ, Olsen HS, Moore PA, Wei Y, Greene JM;
PI Ruben SM, Liu D, Crocker PR;
XX
DR WPI; 2003-567103/53.
DR N-PSDB; ADA27044.
XX
PT New human secreted nucleic acid molecules and polypeptides, useful for
PT preventing, treating, or ameliorating a medical condition, such as
PT cancer, inflammation, immune disorders, neurological and blood clotting
PT disorders.
XX
PS Claim 11; Page 302-305; 454pp; English.

XX The invention relates to an isolated nucleic molecule that is at least
CC 95% identical to 18 human cDNA sequences representing 12 novel genes
CC encoding secreted proteins or a polynucleotide fragment of the cDNA
CC sequence contained in American Type Culture Collection (ATCC) deposit No.
CC defined in the specification, its species homologue, a variant or allelic
CC variant of the polynucleotide having a polynucleotide capable of
CC hybridising under stringent conditions to a nucleic acid molecule
CC does not hybridise under stringent conditions to a nucleic acid molecule
CC having a nucleotide sequence of only A or T residues. Also included are
CC recombinant vectors, host cells (for producing the polypeptide), the
CC secreted polypeptide (comprising a sequence that is at least 95%
CC identical to a polypeptide fragment, domain, epitope, full-length
CC protein, variant, allelic variant or species homologue), antibodies that
CC specifically bind to the polypeptides, diagnosing, treating, preventing
CC or ameliorating a medical condition by administering the polynucleotide
CC or the polypeptide, the gene corresponding to the cDNA sequence and
CC identifying an activity in a biological assay (by expressing the cDNA
CC sequence in a cell, isolating the supernatant, and detecting an activity
CC in a biological assay and identifying the protein in the supernatant
CC having the activity). The polypeptides, nucleic acids and antibodies are
CC useful for diagnosing a pathological condition or a susceptibility to a
CC pathological condition, for preventing, treating, or ameliorating a
CC medical condition, such as cancer, inflammation and other immune
CC disorders, neurological and blood clotting disorders (many examples are
CC given in the specification). The nucleic acids are also useful for
CC chromosome identification, radiation hybrid mapping or long-range
CC restriction mapping. The polypeptides and antibodies are useful for
CC providing immunological probes for differential identification of the
CC tissues immunohistochemistry assays. The polypeptide, polynucleotide,
CC agonist or antagonist may also be used as a food additive or preservative
CC to increase or decrease storage capabilities, fat content or other
CC nutritional components. The present is a secreted protein of the
XX invention.

XX Sequence 1034 AA;
SQ

Query Match 100.0%; Score 1022; DB 6; Length 1034;
Best Local Similarity 100.0%; Pred. No. 1.4e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQTYMDIVILVLDGNSIYPWVEVQHFLINILKFKYIGPGQIQVGWYQGVGVVHFEHLND 60
DB 159 CQTYMDIVILVLDGNSIYPWVEVQHFLINILKFKYIGPGQIQVGWYQGVGVVHFEHLND 218

QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGHSDSP 120
DB 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGHSDSP 278

QY 121 DLEKVIQOSERDNTVRYAVAVLGYNNRRGINPEFLNEIKYIASDPDDKHFNVTDAAAL 180
DB 279 DLEKVIQOSERDNTVRYAVAVLGYNNRRGINPEFLNEIKYIASDPDDKHFNVTDAAAL 338

QY 181 KDIVDALGDRIFSLEGT 197
DB 339 KDIVDALGDRIFSLEGT 355

```

RESULT 6
ADE86592
ID ADE86592 standard; protein; 1034 AA.
XX
XX
AC ADE86592;
XX
XX
DT 29-JAN-2004 (first entry)
XX
XX
DE Novel human secreted protein #15.
XX
XX
KW human; secreted protein; cancer; liver disorder; hepatitis;
KW neural disorder; Alzheimer's disease.
XX
XX
OS Homo sapiens.
XX
XX
FN US2003129685-A1.
XX
XX
PD 10-JUL-2003.
XX
XX
PF 18-APR-2001; 2001US-00836353.
XX
XX
PR 28-OCT-1998; 98US-0105971P.
PR 27-OCT-1999; 99WO-US025031.
PR 19-APR-2000; 2000US-0198407P.
XX
XX
PA (NIJJ/) NI J.
PA (YOUN/) YOUNG P E.
PA (KENN/) KENNY J J.
PA (OLSE/) OLSEN H S.
PA (MOOR/) MOORE P A.
PA (WEIY/) WEI Y.
PA (GREE/) GREENE J M.
PA (RUBE/) RUBEN S M.
XX
XX
PI Ni J, Young PE, Kenny JJ, Olsen HS, Moore PA, Wei Y, Greene JM;
PI Ruben SM;
XX
XX
DR N-PSDB; ADE86574.
XX
XX
PT New nucleic acid molecule, useful for preparing a medicament for
PT preventing, treating or ameliorating a medical condition e.g. cancer,
PT liver disorders or neural disorders.
XX
XX
PS Claim 11; SEQ ID NO 43; 380pp; English.
XX
XX
CC The invention relates to an isolated nucleic acid sequence, or its
CC allelic variant, a fragment of the cDNA sequence, or its fragment,
CC domain, epitope or species homologue. The nucleic acid is useful for
CC preparing a medicament for preventing, treating or ameliorating a medical
CC condition e.g., cancer, liver disorders such as hepatitis or neural
CC disorders such as Alzheimer's disease. The present sequence represents
CC the amino acid sequence of a novel human secreted protein.
XX
XX
SQ Sequence 1034 AA;

Query Match
Best Local Similarity 100.0%; Score 1022; DB 8; Length 1034;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQTYMDIVLDGNSIYPWEVQHFILNLKFKFYIGPGQIQGVVQYQGEDVVHREHLND 60
DB 159 CQTYMDIVLDGNSIYPWEVQHFILNLKFKFYIGPGQIQGVVQYQGEDVVHREHLND 218
QY 61 YRSVKDVEAASHIEQGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGESHDSF 120
DB 219 YRSVKDVEAASHIEQGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGESHDSF 278
QY 121 DLEKVIQSERDNRVTRYAVAVLGYNNRGINPETFLEIKYIASDPPDDKHFFNVDEAAL 180
DB 279 DLEKVIQSERDNRVTRYAVAVLGYNNRGINPETFLEIKYIASDPPDDKHFFNVDEAAL 338

```

```

QY 181 KDIVDALGDRIFSLEGT 197
DB 339 KDIVDALGDRIFSLEGT 355

RESULT 7
ABR58365
ID ABR58365 standard; protein; 1120 AA.
XX
XX
AC ABR58365;
XX
XX
DT 07-JUL-2003 (first entry)
XX
XX
DE Human NOV2b.
XX
XX
KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;
KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;
KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;
KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;
KW neurodegenerative disorder; Alzheimer's disease; immune disorder;
KW haematopoietic disorder.
XX
XX
OS Homo sapiens.
XX
XX
FN WO2003029423-A2.
XX
XX
PD 10-APR-2003.
XX
XX
PF 02-OCT-2002; 2002WO-US031358.
XX
XX
PR 02-OCT-2001; 2001US-0326483P.
PR 05-OCT-2001; 2001US-0327342P.
PR 09-OCT-2001; 2001US-0327917P.
PR 09-OCT-2001; 2001US-0328029P.
PR 09-OCT-2001; 2001US-0328044P.
PR 09-OCT-2001; 2001US-0328056P.
PR 12-OCT-2001; 2001US-0328849P.
PR 15-OCT-2001; 2001US-0329414P.
PR 17-OCT-2001; 2001US-0330142P.
PR 22-OCT-2001; 2001US-0341058P.
PR 24-OCT-2001; 2001US-0339266P.
PR 24-OCT-2001; 2001US-0343629P.
PR 29-OCT-2001; 2001US-0349575P.
PR 01-NOV-2001; 2001US-0346357P.
PR 12-APR-2002; 2002US-0371972P.
PR 12-APR-2002; 2002US-0371980P.
PR 17-APR-2002; 2002US-0373261P.
PR 19-APR-2002; 2002US-0373805P.
PR 23-APR-2002; 2002US-0374738P.
PR 16-MAY-2002; 2002US-0381101P.
PR 17-MAY-2002; 2002US-0381635P.
PR 29-MAY-2002; 2002US-0383830P.
PR 01-OCT-2002; 2002US-00262839.
XX
XX
PA (CURA-) CURAGEN CORP.
XX
XX
PI Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
PI Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;
XX
XX
DR WPI; 2003-381625/36.
DR N-PSDB; ACC72077.
XX
XX
PT NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
PT treating NOVX associated disorders, e.g. diabetes, obesity, cancer or
PT dyslipidaemia, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX
PS Claim 1; Page 107; 487pp; English.
XX

```

CC The present invention relates to novel human NOV proteins and their
CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV
CC proteins are useful in manufacturing a medicament for treating a syndrome
CC associated with a human disease. The NOV proteins and coding sequences
CC may be used to diagnose, treat or prevent metabolic disorders such as
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative
CC disorders such as Alzheimer's disease or Parkinson's disease, immune
CC disorders, haematopoietic disorders and various dyslipidaemias
XX
SQ Sequence 1120 AA;

Query Match 100.0%; Score 1022; DB 6; Length 1120;
Best Local Similarity 100.0%; Pred. NO. 1.6e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQTYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGGQIQGVGVQGEDVWVEHFLND 60
DB 90 CQTYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGGQIQGVGVQGEDVWVEHFLND 149
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKGGKAKKVMIVITDGHSDSP 120
DB 150 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKGGKAKKVMIVITDGHSDSP 209
QY 121 DLEKVIQOSERDNTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 180
DB 210 DLEKVIQOSERDNTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 269
QY 181 KDIVDALGDRIFSLEGT 197
DB 270 KDIVDALGDRIFSLEGT 286

RESULT 8
AAB30929
ID AAB30929 standard; protein; 1188 AA.

XX AAB30929;
XX
DT 02-APR-2001 (first entry)
XX

DE Amino acid sequence of a human alpha11 integrin chain.
XX Human; integrin; alpha11 subunit; fibroblast; muscle cell; chondrocyte;
XX osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;
KW wound healing; trauma; rheumatoid arthritis; osteoarthritis;
KW osteoporosis; cartilage damage; bone damage; cartilage.
XX

OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /note= "signal peptide"
FT Region 951..972
FT /note= "leucine zipper"
FT Domain 1142..1164
FT /note= "transmembrane domain"

XX WO200075187-A1.
XX PN
XX PD 14-DEC-2000.
XX PF 31-MAY-2000; 2000WO-SE001135.
XX PR 03-JUN-1999; 99SE-00002056.
XX PA (ACTI-) ACTIVE BIOTECH AB.
XX PI Gullberg D;
XX WPI; 2001-071061/08.
DR N-PSDB; AAC86871.
XX Integrin subunit alpha 11 or integrin heterodimer comprising subunit
PT

PT alpha 11 in association with subunit beta, useful for treating muscle
PT dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis.
XX
PS Disclosure; Fig 2a-c; 79pp; English.

XX The present sequence represents a human integrin subunit, designated
CC alpha11. The alpha11 polynucleotide and polypeptide are useful as markers
CC of cell target molecules, such as fibroblasts, muscle cells,
CC chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.
CC They are also used for determining the differential-stage of cells during
CC differentiation, development in pathological conditions, in tissue
CC regeneration, in transplantation or in therapeutic and physiological
CC repair of tissues. The pathological conditions involving subunit alpha11
CC are selected from damage of cells, muscle dystrophy, fibrosis, wound
CC healing, trauma, rheumatoid arthritis, osteoarthritis and osteoporosis,
CC damage of cartilage and bone, and cartilage and bone diseases. The
CC polypeptide is useful for detecting the formation of cartilage during
CC embryonic development, for detecting physiological therapeutic repair of
CC cartilage and muscle, for selection and analysis, or for sorting,
CC isolating or purification of chondrocytes and muscle cells, for detecting
CC regeneration of cartilage or chondrocytes during transplantation of
CC cartilage or chondrocytes during transplantation of cartilage or
CC chondrocytes, respectively, or of muscle or muscle cells during
CC transplantation of muscle or muscle cells, respectively, and for studies
CC of differentiation or chondrocytes or muscle cells
XX
SQ Sequence 1188 AA;

Query Match 100.0%; Score 1022; DB 4; Length 1188;
Best Local Similarity 100.0%; Pred. NO. 1.7e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQTYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGGQIQGVGVQGEDVWVEHFLND 60
DB 159 CQTYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGGQIQGVGVQGEDVWVEHFLND 218
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKGGKAKKVMIVITDGHSDSP 120
DB 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKGGKAKKVMIVITDGHSDSP 278
QY 121 DLEKVIQOSERDNTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 180
DB 279 DLEKVIQOSERDNTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
DB 339 KDIVDALGDRIFSLEGT 355

RESULT 9
AAU14467
ID AAU14467 standard; protein; 1188 AA.
XX AAU14467;
XX AC AAU14467;
XX DT 24-OCT-2001 (first entry)
XX DE Human novel protein #338.
XX
XX Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; Gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
XX OS Homo sapiens.
XX PN WO200155437-A2.
XX XX
XX 02-AUG-2001.
XX

```
PF 25-JAN-2001; 2001WO-US002623.
XX
XX
PR 25-JAN-2000; 2000US-00491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
XX
DR N-PSDB; AAS22772.
XX
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage.
XX
XX
PS Example 4; Page 828-831; 894pp; English.
XX
XX
CC The invention relates to polynucleotides encoding novel human proteins or
CC their active domains. The polypeptides, polynucleotides and antibodies
CC raised against the polypeptides are used in a method of treatment of a
CC mammal and prevention of disorders caused by the aberrant protein
CC expression or activity. The polypeptides can be used as molecular weight
CC markers, food supplements, and in antibody production. The polypeptides
CC are used to identify compounds which bind to the polypeptides.
CC Polynucleotides of the invention are used as probes and primers, for
CC sequencing, for chromosome or gene mapping, in the production of
CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
CC therapy. Polypeptides of the invention can be used to target drugs to a
CC tumour, in assays to determine biological activity, to raise
CC antibodies/ elicit an immune response, to determine quantitative protein
CC levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
CC diseases, nervous system disorders, and infection. The present sequence
XX represents a protein of the invention
XX
SQ Sequence 1188 AA;
Query Match 100.0%; Score 1022; DB 4; Length 1188;
Best Local Similarity 100.0%; Pred. No. 1.7e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CQTYMDIVILDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVWVHEFLND 60
DB 159 CQTYMDIVILDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVWVHEFLND 218
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSHP 120
DB 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSHP 278
QY 121 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFNVTDEAL 180
DB 279 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFNVTDEAL 338
QY 181 KDIVDALGRIFSLEGT 197
DB 339 KDIVDALGRIFSLEGT 355
RESULT 10
AAU14231
XX AAU14231 standard; protein; 1188 AA.
XX AC AAU14231;
XX DT 24-OCT-2001 (first entry)
XX
```

```
DE Human novel protein #102.
XX
XX Human, novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
XX WO2000155437-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 25-JAN-2001; 2001WO-US002623.
PF
XX
XX 25-JAN-2000; 2000US-00491404.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-451939/48.
DR
XX
XX N-PSDB; AAS22536.
DR
XX
XX Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage.
XX
XX Example 4; Page 578-581; 894pp; English.
XX
XX The invention relates to polynucleotides encoding novel human proteins or
XX their active domains. The polypeptides, polynucleotides and antibodies
XX raised against the polypeptides are used in a method of treatment of a
XX mammal and prevention of disorders caused by the aberrant protein
XX expression or activity. The polypeptides can be used as molecular weight
XX markers, food supplements, and in antibody production. The polypeptides
XX are used to identify compounds which bind to the polypeptides.
XX Polynucleotides of the invention are used as probes and primers, for
XX sequencing, for chromosome or gene mapping, in the production of
XX recombinant proteins, and in generating anti-sense DNA or RNA and in gene
XX therapy. Polypeptides of the invention can be used to target drugs to a
XX tumour, in assays to determine biological activity, to raise
XX antibodies/ elicit an immune response, to determine quantitative protein
XX levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
XX versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
XX diseases, nervous system disorders, and infection. The present sequence
XX represents a protein of the invention
XX
SQ Sequence 1188 AA;
Query Match 100.0%; Score 1022; DB 4; Length 1188;
Best Local Similarity 100.0%; Pred. No. 1.7e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CQTYMDIVILDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVWVHEFLND 60
DB 159 CQTYMDIVILDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVWVHEFLND 218
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSHP 120
DB 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSHP 278
QY 121 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFNVTDEAL 180
DB 279 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFNVTDEAL 338
QY 181 KDIVDALGRIFSLEGT 197
DB 339 KDIVDALGRIFSLEGT 355
```

Db 279 DLEKVIQQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDDPKHFFNVTDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
Db 339 KDIVDALGDRIFSLEGT 355
RESULT 11
AAB50085
ID AAB50085 standard; protein; 1188 AA.
XX
AC AAB50085;
XX
DT 19-MAR-2001 (first entry)
XX
DE Human A259.
XX
KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
KW rheumatoid arthritis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..1141
FT /label= Extracellular_domain
FT Peptide 1..22
FT /label= Signal_peptide
FT Protein 23..1188
FT /label= Mature_protein
FT Domain 39..74
FT /label= Integrin_alphasubunit_repeat_domain_#1
FT Domain 115..157
FT /label= Integrin_alphasubunit_repeat_domain_#2
FT Domain 164..345
FT /label= I_domain
FT Domain 367..392
FT /label= Integrin_alphasubunit_repeat_domain_#3
FT Domain 421..455
FT /label= Integrin_alphasubunit_repeat_domain_#4
FT Domain 478..516
FT /label= Integrin_alphasubunit_repeat_domain_#5
FT Domain 540..575
FT /label= Integrin_alphasubunit_repeat_domain_#6
FT Domain 602..640
FT /label= Integrin_alphasubunit_repeat_domain_#7
FT Domain 1142..1164
FT /label= Transmembrane_domain
FT Domain 1165..1188
FT /label= Cytoplasmic_domain
XX
PN WO200073339-A1.
XX
PD 07-DEC-2000.
XX
PF 15-MAY-2000; 2000WO-US013262.
XX
PR 28-MAY-1999; 99US-00322790.
PR 27-APR-2000; 2000US-00561263.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Pan Y, Lora JM;
XX
DR WPI; 2001-041142/05.
DR N-PSDB; AAC91901, AAC91902.
XX
XX Nucleic acid encoding alpha-integrin subunits, useful for treatment and
PT diagnosis of fibrosis, e.g. of the liver.
XX
PS Claim 8; Fig 1; 164pp; English.
XX

CC The present sequence is human integrin alpha subunit, A259. A259 is
CC homologous with the alpha1 and alpha10 integrin subunits and is
CC overexpressed in fibrosis. A259 is implicated in regulation of
CC proliferation, differentiation and/or function of many different cell
CC types. Inhibitors of A259 activity are useful for the treatment of liver
CC disease, particularly fibrosis, and also fibrosis in other organs
CC (specifically lung and kidney). In addition, A259 can be used for
CC treatment and prevention of cancer, osteoporosis, acute myeloid
CC leukaemia, HIV infection, and rheumatoid arthritis
XX
SQ Sequence 1188 AA;
Query Match 100.0%; Score 1022; DB 4; Length 1188;
Best Local Similarity 100.0%; Pred. No. 1.7e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CQTYMDIVILDGSNSIYPWVEVQHFLINILKFFYIGPGQIQGVVQYGEDVVHPEHLND 60
Db 159 CQTYMDIVILDGSNSIYPWVEVQHFLINILKFFYIGPGQIQGVVQYGEDVVHPEHLND 218
QY 61 YRSVKDWEAASHIEORGTTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDSP 120
Db 219 YRSVKDWEAASHIEORGTTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDSP 278
QY 121 DLEKVIQQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDDPKHFFNVTDEAAL 180
Db 279 DLEKVIQQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDDPKHFFNVTDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
Db 339 KDIVDALGDRIFSLEGT 355
RESULT 12
AAB50551
ID AAB50551 standard; protein; 1188 AA.
XX
AC AAB50551;
XX
DT 14-FEB-2002 (first entry)
XX
DE Human A259 polypeptide.
XX
KW Human; A259; integrin alpha subunit; integrin alpha 10; secreted protein;
KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
KW cartilage associated disorder; haematopoietic disorder; bone marrow;
KW immune related disease; apoptotic disorder; neuronal tissue disease;
KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
KW antiarthritic; antianaemic; antiallergic; antiasthmatic; dermatological;
KW antidiabetic; anticonvulsant; antiparkinsonian.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..1141
FT /note= "Extracellular domain"
FT Peptide 1..22
FT /note= "Signal peptide"
FT Protein 23..1188
FT /note= "Mature human A259"
FT Domain 37..90
FT /note= "Integrin alpha repeat domain. The specification
FT states that this domain exists in human A549"
FT Domain 115..157
FT /note= "Integrin alpha repeat domain"
FT Domain 164..345
FT /note= "I domain or Von Willebrand Factor type A domain"
FT Domain 367..392
FT /note= "Integrin alpha repeat domain"
FT Domain 421..472
FT /note= "Integrin alpha repeat domain. The specification
FT states that this domain exists in human A549"
FT

FT Domain 476. .532
 FT /note= "Integrin alpha repeat domain. The specification
 FT states that this domain exists in human A549"
 FT 538. .593
 FT Domain
 FT /note= "Integrin alpha repeat domain. The specification
 FT states that this domain exists in human A549"
 FT 600. .654
 FT /note= "Integrin alpha repeat domain. The specification
 FT states that this domain exists in human A549"
 FT 1142. .1164
 FT Domain
 FT /note= "Transmembrane domain"
 FT 1165. .1188
 FT /note= "Cytoplasmic domain"
 FT
 XX
 PN WO200181414-A2.
 XX
 XX
 PD 01-NOV-2001.
 XX
 XX 27-APR-2001; 2001WO-US013516.
 XX
 XX 27-APR-2000; 2000US-00561263.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 XX Pan Y, Lora J;
 XX
 XX WPI; 2002-041397/05.
 DR N-PSDB; AAS16873.
 DR
 XX
 XX New A259 nucleic acids and polypeptides, which comprise integrin alpha
 FT subunit, useful for diagnosing, preventing or treating e.g. liver
 PT disease, kidney or lung fibrosis, cancers, blood disorders or immune
 PT related diseases.
 XX
 PS Claim 9; Fig 1; 168pp; English.
 XX
 CC The invention relates to human and murine A259 nucleic acid molecules
 CC which encode secreted proteins with homology to integrin alpha subunits,
 CC specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
 CC are useful for treating liver disease or fibrosis, particularly kidney
 CC fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
 CC useful for diagnosing, preventing or treating cartilage and bone
 CC associated disorders (such as bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
 CC osteoporosis), bone marrow, blood and haematopoietic disorders (such as
 CC acute myeloid leukaemia, haemophilia, anaemia and thalassaemia), immune
 CC related diseases (such as HIV, viral infections, cancers, T cell
 CC autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
 CC asthma and psoriasis), apoptotic disorders (such as systemic lupus
 CC erythematosus and insulin-dependent diabetes mellitus), diseases of the
 CC neuronal tissues (such as epilepsy and muscular dystrophy) and
 CC neurodegenerative diseases (such as Parkinson's disease and Huntington's
 CC disease). This sequence represents the human A259 polypeptide
 XX
 SQ Sequence 1188 AA;
 Query Match 100.0%; Score 1022; DB 5; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 1.7e-98;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CQTYMDIVLDGNSIYPWVEVQHFLINLKFKFYIGPGQIQGVVQYGEDVVHFLND 60
 DB 159 CQTYMDIVLDGNSIYPWVEVQHFLINLKFKFYIGPGQIQGVVQYGEDVVHFLND 218
 QY 61 YRSVKDVEAASHIEORGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSNDSP 120
 DB 219 YRSVKDVEAASHIEORGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSNDSP 278
 QY 121 DLEKVIQOSERNVTRYAVAVILGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL 180
 DB 279 DLEKVIQOSERNVTRYAVAVILGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL 338
 QY 181 KDIVDALGDRIFSLEGT 197

Db 339 KDIVDALGDRIFSLEGT 355
 RESULT 13
 ADE09956
 ID ADE09956 standard; protein; 1188 AA.
 XX
 AC ADE09956;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Novel protein-related contig polypeptide sequence #544.
 XX
 KW novel gene; novel protein; tissue marker; molecular weight marker;
 KW chromosome marker; genetic disorder; contig.
 XX
 OS Unidentified.
 XX
 PN WO2003054152-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 10-DEC-2002; 2002WO-US039555.
 XX
 XX 10-DEC-2001; 2001US-0339739P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.
 PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-037381P.
 PR 12-APR-2002; 2002US-0373615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou F, Dimanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
 XX WPI; 2003-569235/53.
 DR
 XX New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 PS Disclosure; SEQ ID NO 3022; 1177pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence was used in the
 CC exemplification of the invention.
 XX
 SQ Sequence 1188 AA;
 Query Match 100.0%; Score 1022; DB 7; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 1.7e-98;
 Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CQTYMDIVLDGNSIYPWVEVQHFLINLKFKFYIGPGQIQGVVQYGEDVVHFLND 60
 DB 159 CQTYMDIVLDGNSIYPWVEVQHFLINLKFKFYIGPGQIQGVVQYGEDVVHFLND 218
 QY 61 YRSVKDVEAASHIEORGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSNDSP 120
 DB 219 YRSVKDVEAASHIEORGGTETRTAFGIEFARSEAFQKGRGAKKVMIVITDGSNDSP 278
 QY 121 DLEKVIQOSERNVTRYAVAVILGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL 180

Db 279 DLEKVIQSQERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
Db 339 KDIVDALGDRIFSLEGT 355
RESULT 14
AA25582
ID AAB25582 standard; protein; 1189 AA.
XX
AC AAB25582;
XX
DT 21-NOV-2000 (first entry)
XX
DE ITGA11 protein encoded by human secreted protein gene #7.
XX
KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
KW antihematic; dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnery; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
XX
OS Homo sapiens.
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX 27-OCT-1999; 99WO-US025031.
XX
XX 28-OCT-1998; 98US-0105971P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
XX Greene JW;
XX
XX WPI; 2000-387742/33.
XX
XX N-PSDB; AAA80612.
XX
XX Isolated nucleic acid molecules encoding human secreted proteins are used
XX for the prevention, amelioration and treatment of autoimmune,
XX PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
XX PT wounds, and infectious diseases.
XX
XX
XX Claim 1; Fig 19A-F; 803pp; English.
XX
XX The present invention relates to 12 secreted human proteins and the
XX nucleotide sequences encoding them. The polynucleotide sequences given in
XX AAA80606-A80623 encode the 12 secreted protein sequences given in
XX AAB25576-B25593. The human secreted proteins have various activities
XX dependent on the tissues in which they are expressed. Examples of the
XX activities of the proteins include: immunosuppressant; anti-inflammatory;
XX antiarthritic; antirheumatic; dermatological; antiproliferative;
XX antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial;
XX and antifungal activity. The proteins, polypeptides, agonists and
XX antagonists may be used to treat prevent and/or diagnose various disease,
XX disorders and conditions examples of which include: immune disorders e.g.
XX Addison's disease, rheumatoid arthritis, dermatitis, and multiple
XX sclerosis; inflammatory disorders e.g. inflammatory bowel disease.
XX Crohn's disease and nephritis; hyperproliferative disorders such as
XX paraproteinemia and purpura; cardiovascular disorders e.g. coronary
XX arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The
XX proteins and polynucleotide sequences may also be used in wound healing
XX and the treatment of infectious diseases. The human secreted protein gene
XX #7 and protein sequences are represented in sequences AAA80612 and
XX AAB25582. Secreted protein gene #7 is located at position chromosome 15
XX q22.3-23. Sequences AAA80612-A80661 represent genes which are related to
XX the secreted protein gene#7

XX
SQ Sequence 1189 AA;
Query Match 100.0%; Score 1022; DB 3; Length 1189;
Best Local Similarity 100.0%; Pred. No. 1.7e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CQTYMDIVIVLDSNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGDVVVHEPHLND 60
Db 159 CQTYMDIVIVLDSNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGDVVVHEPHLND 218
QY 61 YRSVKDWEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDSP 120
Db 219 YRSVKDWEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDSP 278
QY 121 DLEKVIQSQERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180
Db 279 DLEKVIQSQERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
Db 339 KDIVDALGDRIFSLEGT 355
RESULT 15
ABG12949
ID ABG12949 standard; protein; 1189 AA.
XX
AC ABG12949;
XX
DT 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #12940.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS77136.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 43308; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 1189 AA;

```
Query Match      100.0%; Score 1022; DB 4; Length 1189;
Best Local Similarity 100.0%; Pred. No. 1.7e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CQTYMDIVIVLDGNSIYPWVEVQHFLNLIKFKYIGPGQIQVGVOYGEDVWHEFHLND 60
Db      159 CQTYMDIVIVLDGNSIYPWVEVQHFLNLIKFKYIGPGQIQVGVOYGEDVWHEFHLND 218

QY      61 YRSVKDVVEAAASHIEQGGTETRTAFGIEFARSEAFQGGKKGAKKVMIVITDGESHDSP 120
Db      219 YRSVKDVVEAAASHIEQGGTETRTAFGIEFARSEAFQGGKKGAKKVMIVITDGESHDSP 278

QY      121 DLEKVIQQSERDQVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL 180
Db      279 DLEKVIQQSERDQVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL 338

QY      181 KDIVDALGDRIFSLEGT 197
Db      339 KDIVDALGDRIFSLEGT 355
```

Search completed: September 21, 2004, 12:59:16
Job time : 38.5566 secs

This Page Blank (uspto)

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2159	34.7	1151	2	A45226	integrin alpha-1 c	
2	2149.5	34.5	1180	2	A35854	integrin alpha-1 c	
3	1910.5	30.7	1178	2	S44142	VLA-2 protein homol	
4	1901.5	30.6	1170	2	I45914	integrin alpha-2 s	
5	1863	29.9	1181	2	A33998	integrin alpha-2 c	
6	1146.5	18.4	1170	2	S03308	cell surface glyco	
7	1135.5	18.2	1163	1	RWHU1C	cell surface glyco	
8	1110.4	17.7	1153	2	S00551	leukocyte surface	
9	1094	17.6	1153	1	RWHU1B	cell surface glyco	
10	1073.5	17.2	1163	2	I56126	lymphocyte function	
11	924.5	14.9	1179	2	A53213	integrin alpha-E c	
12	799.5	12.8	1035	2	I58409	integrin alpha-9 c	
13	789	12.7	272	2	A53448	integrin alpha-1 -	
14	738	11.9	1041	2	T31437	integrin alpha cha	
15	719.5	11.6	1054	2	JC7294	alphan integrin -	
16	693.5	11.1	1039	2	A41131	lymphocyte-Peyer's	
17	672.5	10.8	1038	2	S06046	integrin alpha-4 c	
18	607.5	9.8	1034	2	A36108	integrin alpha-V c	
19	606	9.7	1049	2	A27109	integrin alpha-4 c	
20	605	9.7	1091	2	A41543	fibronectin recept	
21	604.5	9.7	1053	2	S44250	integrin alpha-6 c	
22	604	9.7	1051	2	A40021	integrin alpha-5 c	
23	602.5	9.7	1073	2	B36429	integrin alpha-3 c	
24	594	9.5	1053	2	I55534	integrin alpha-6 c	
25	587	9.4	1051	2	A35761	integrin alpha-6 c	
26	584.5	9.4	1044	2	T10050	VLA-3 alpha subuni	
27	568	9.1	1072	2	A38457	cell surface glyco	
28	568.5	9.1	1072	2	A38457	integrin alpha-v c	
29	563	9.0	1146	2	S40311	integrin alpha-6 c	
30	563	9.0	1044	2	S03516	integrin - fruit f	
31	563	9.0	1044	2	S03516	integrin alpha-8 c	

Db	419	SGDVLYIAGQPRYNHTQVILYRMEDG-NIKILOQLSGEQISYFGSILTTTDDKDSNT	477	Qy	11	WALSILWPGFTDTFNMTRKPRVIPSRTAFPGYTVQQHDIISGNKWLVGAPLETTYQKT	70
Qy	496	DVLLVGAQMYF-NEGREGKVVVYELRQNRFFVINGTLK-----DSHSYQN---	539	Db	17	WLLTVILGFCVFNVDVKNSMSFGPVEDMEGYTVQOYENEEGKWVLIGSLVGPQKART	76
Db	478	DILLVGAQMYGTEKEEGKVVYVAINLTREYQMSLEPIKQTCSSRQHSCTTENKNE	537	Qy	71	GDVYKCPVIGHN---CTKLNIG-RVTLGNSYERKDNMLGLSLATNPKDNFGLACSPWS	126
Qy	540	---ARFGSSIASVRDLNQSDVNDVVVGAPLEDNHAGAIYIHFGRGSILKTPKQITASE	596	Db	77	GDVYKCPVGRERAMPCKVLDLPVNTSIPNVTEIKENMTFTGLVTNP--NGGFLACGPLYA	135
Db	538	PCGARFGTAAVAKOLNLDGENDIVIGAPLEDHGGAVIVYHSGKTIKKEYAQRIPSGG	597	Qy	127	HECGSSYTTGCMGRVNSNFRFSKTVAPALQCCQYMDIVIVLDGNSISYPMWVEOHLI	186
Qy	597	LATGLOYGCSHGQIDLNEDGLIDLAVGALGNVILMSRPVQVNASLHEEPSKINIFH	656	Db	136	YRCGHLHYTTGICSDVSPTFQVNSFAP-VQECSTQLDIVIVLDGNSISYPMWVEOHLI	194
Db	598	DGKTLKFTGQSHGEMDLDGSLTDTVTGGIGGAALFWSRDVAVVKVWNFPFNKVNQK	657	Qy	167	NILKKFYIGPGIOGVVQYGEDVVVHEHLADYRSKDVVAAASHIEORGGTETATAGI	246
Qy	657	ROCKSGSDATCLAAFLCFTPIFLAPHFQTTVWIRYNATMDRERYTPRAHLDGEGDRFT	716	Db	195	DLLKRMDLGPQQTQVGIQVYGENVTHEFNLNKYSSTEEVVAANKIGRGGQLTMTALGI	254
Db	658	KNCHMEGKETVCINATVCEVVKLSKEDPIYENDLQYRVTLDSLRQISRSFSGTOERKV	717	Qy	247	EFARSEAF--QKGGKSGAKKVMIVITDGESHSDSPLEKVIQOESRDNVTRAVAVLGYN	304
Qy	717	NRAVLSSQELCERINFHVLDTADYVKPVTFSVEYSLEDPDHGPMDDCMTTLRVSP	776	Db	255	DTARKEAFTEARGARGGVKVMVIVITDGESHNDYRLKQVQDCEDENTQRFSAILGHYN	314
Db	718	QRNITVRKSE--CTKHSFYMLDKHDFQDSVRITLDFNLTPDNGFPVLDLSPNSVHEYIP	775	Qy	305	RGINPETFLEIKYIASDPDDKHFFENVTDAAKXIDVADLGDRIFFSLGTT-NKNETSFG	363
Qy	777	FWNGCNEDEHCVDPDLVDARSDLPTAMEYQORVLRKPAQDCSAYTLSPDTPVFIESTRQ	836	Db	315	RGNLTKEFVEBIKSIASEPTEKHFFNVSDLEALVTIVKALGERIFALEATADQSAASE	374
Db	776	FAKDCGNKEKCSIDL-----SLHVATTEKDLILYRSQND	809	Qy	364	LEMSOTGFSHVVEDGVILGAVGAYDMGAVLKETSAGKVIPLRSYLYKEFPEELKNHGA	423
Qy	837	RVAVEATLENRGENAYSTVLNISQSANLOFASI--IQKEDSDGSECVNEERRLOKQVCN	894	Db	375	MEMSOTGSAHYSQDWMLGAVGADWNGTVVMQXANQWVIPHNTTFTQEPAKMNEPLAS	434
Db	810	KFNVSITVKNTKDSANTRTIVHYSPLNLFSGIEALQKDCESN-----HNITCK	859	Qy	424	YLGTYTVTSVSSROGRV-YVAGAPRNTGKVLFTMNNRSLTTHQAMRGQOIGSYFSG	482
Qy	895	VSYPPFAKAKVAFRLDSFESKIFLHLEIELAGSDSNERNSTKEDNVAPLRFHLKYE	954	Db	435	YLGTYWNS--ATIPGVLYIAGQPRYNHTGQVVIYKMDG-NINILQTLGGEQIGSYFSG	491
Db	860	VGYPFLRRGEMVTKILFQNTSYLMENVTYILSATSEEPETLSDNVNIAISIPVKYE	919	Qy	483	EITSVDIDGDGVTVLLVGAQMYF-NEGREGKVVVYELRQNRFFVN-----GT	530
Qy	955	ADVLPFRSSLSHVEVKLNS-----LERYDGIQPPSCIFRIQNLGLFPHGMMKITTIP	1010	Db	492	VLTITDIDKDSYTDLLVGAPMYGTEKEEGKVVYVAVNQTRFQMSLEPIRQTCSSS	551
Db	920	VGLQFYSSASEYHISAAETVPEVINSTEDIGNEINIFYLRKSSFPMPPELKLSISFP	979	Qy	531	LKDSHSYQ-----NAREGSSIASVRDLNQSDVNDVVVGAPLEDNHAGAIYIHFGRGS	583
Qy	1011	IATFRGNRLIKLDFLTDEANTSC-----NIMGNSTEYRPTVPEBDRAPOLNHSNSD	1064	Db	552	LKDNSTCKENKNEPCGARFGTAAVKDLNVDFGNDVVGAPLEDDHAGAVIYHSGKT	611
Db	980	NMTSNGYPVL-YPTGLSSSENANCRPHIFEDPPSINSKQMTTSTDLKRGTLIDCNTCK	1038	Qy	584	ILKTPQRTITABELATGLQYFGCSHGQILDLEDGLIDLAVGALGNVILMSRPVQVINA	643
Qy	1065	VSGINCNIRLVNQENFHLGNLWLSLKALKYKMKIMVNAALQORQHSPIFREEDP	1124	Db	612	IREAVAQRIPSGGDGKTLKFFGOSIHGEMDLNGDGLTDTVTIGLGGAAALFWARDVAVKV	671
Db	1039	FATITCNLTSSDISQVNVSLI--LWKPTFIKSYFSSNLITIRGEL-RSENASVLSSNSQ	1095	Qy	644	SLHFESKINIHRDCKRSRDATCLAAFLCFTPIFLAPHFQTTVWIRYNATMDRERYT	703
Qy	1125	SROIETEISQOE-DWQVPIWIVGSTLGLLALLVLRKLGFFRPSARRR	1176	Db	672	TNFEFENKVINQKNCRVEGKTEVCINATVCHVKLSKEDSIYEADLQYRVTLDSLRQI	731
Db	1096	KRELAIQISKDGLPGRVPLWVILLAPAGULLLLMLLILALWKIGFFKRLPKKK	1148	Qy	704	PRAHLDGEGDRFTNRAVLSSQELCERINFHVLDTADYVKPVTFSVEYSLEDPDHGPM	763
RESULT 2				Db	732	SRSFFSGTQERKIQRNITVRESE--CIRHSFYMLDKHDFQDSVRVTLDFNLTPDNGPVL	789
A35854				Qy	764	DDGWPTTLRVSPFWNGCNEDEHCVDPDLVDARSDLPTAMEYQORVLRKPAQDCSAYT	823
Integrin alpha-1 chain precursor - rat				Db	790	DDALPNSVHEHIPFAKDCGNKERCISDLTLNVSTT-----	824
C;Species: Rattus norvegicus (Norway rat)				Qy	824	FDVTVFIESTRQVAVEATELNRGENAYSTVLNISQSANLOFASI--IQKEDSDGSEIC	881
C;Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 15-Sep-2003				Db	825	-EKSLIVKQHQDKFNVLTKVKNKGSAYNTRIVVQHSPLNIFSGIEEIQKDCESN---	880
C;Accession: A35854; S11243				Qy	882	VNEERLQKQCVNSVPPFPAKAKVAPRLDSEFSKISIFLHLEIELAGSDSNERNSTKE	941
R;Ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto, J. Cell Biol. 111, 709-720, 1990				Db	881	-----QNITCRGVYFPFLRAGETVTFKIIFQENTSHLSENAITHLSATSDSEPLESLN	933
A;Title: Molecular cloning of the rat integrin alpha-1 subunit: a receptor for laminin a				Qy	942	DNVAPLRFHLKYADVLFTRSSLSHYEVKLNLSERY-----DGIGPPFCIERIONLGL	997
A;Accession: A35854				Db	934	DNEVNISIPVKYEVGLQFYSSASEHLSVAAANETIPPEFINSTEDIGNEINVFYTRRGH	993
A;Status: preliminary				Qy	998	FPFHGMMKTIPIATRSGRNLLKLRDLTDEANTSCNIGWNSTEYRPTVEE--DLRRA	1055
A;Molecule type: mRNA				Db	994	FPPELQISLISFPNLITADGVFLYPIGCMSSSD-NVNC-----RPSLEDDPGFINS	1043
A;Residues: 1-1180 <IGN>				Qy	1056	PQLNHSNSDVV-----SINCNIPLVPNQENFPHLLGNLWLSLKALKYKS	1100
A;Cross-references: GB:X52140; NID:G56493; PIDN:CAA36384.1; PID:G56494							
A;Keywords: cell adhesion; cytoskeleton; transmembrane protein							
F;170-345/Domain: von Willebrand factor type A repeat homology <WAA>							
Query Match							
Best Local Similarity							
Matches 463; Conservative 242; Mismatches 405; Indels 107; Gaps 24;							

```

Db 1044 KMTWTSKSEVLKRGITQDCSTCGVATITCSLLPSLSQVNSLL--LWKPIFIRAHFSS 1101
QY 1101 KKMVNAALQROPHSPPIFREDDPSQIEFETSKQE-DWQVPIWIVGTLGGLLLALL 1159
Db 1102 LNLTRGELKSE-NSSLTSLSSNRKRETAIQISKDGLPGRVPLWILLSAFAGLLMLL 1160
QY 1160 VIALRKLGFRRARRR 1176
Db 1161 ILALWKIGFFKRLKKK 1177

RESULT 3
S44142
VLA-2 protein homolog - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Sep-2003
C;Accession: S44142
R;Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latze
submitted to the EMBL Data Library, January 1994
A;Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not
A;Reference number: S44142
A;Accession: S44142
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1178 <EDE>
A;Cross-references: EMBL:Z29987; NID:G473098; PIDN:CAA82877.1; PID:G473099
F;169-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 30.7%; Score 1910.5; DB 2; Length 1178;
Best Local Similarity 36.2%; Pred. No. 3.5e-119;
Matches 442; Conservative 238; Mismatches 437; Indels 103; Gaps 31;

QY 6 GLVVAWALSLPGFTD--TENMDTRKPRVIGSRRTAFGYTVQVQHDISGNKWLVVGAPLE 63
Db 8 GALLQLLMLVQGIINCLAYINVLPGAKIFGSPSEQGYGVQVQ/TNPGQGNLLVGSWS 67
QY 64 TNGYQKTGDVYKCPV--IHGNCITKLN--GRVTLGNVSRKDNMRGLSLATNPKNDSFLA 120
Db 68 GFPENRMGDVYKCPVDLPATACEKLNLSNASISNVTEIKTNMGLTLTRNPGTGGFLT 127
QY 121 CSPLWSHCGSSYTTGMCSRVNPNRFRKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
Db 128 CGPLWAHQCGNQYATGTCSDVSPDFLTFSFPAVQACPSLVDVWVVCDESNIYPWEA 187
QY 181 VOHELINLKKFYIGPGQIOGVVQYGDVHVHEHLNDYRSVKDVVERAASHIEORGGTET 240
Db 188 VKPLVVFVTGLDIPKKTQVALQYANEPRIIFNLNDFETKEDMVQATSETRGHGDLT 247
QY 241 RTAFGIEFARSEAFQK--GGRKGAKVMIVITDGEHSDSPLEKVIQGSERDNNVRYAVA 298
Db 248 NTFRAIEFARDYAYSQTSGGRRPGATKVMVVVTDGESHGSKLTVIQOQNDDEILRFGIA 307
QY 299 VLGYNRRGINPETFELNEIKYIASDDPKHFFNVTDDEALKDVIDALGDRIFSLEGTKN 358
Db 308 VLGLNENALDTKNLKEIKALASTPTERYFFNVFADEAALLEKAGTLGEQIFSEGTGQV 367
QY 359 ETSFGLBMSQTGFSSHVV--EDGVLLGAVGAYDNGAVLKETSGKVIPIRESVLKFPPE 416
Db 368 GDNFQEMAQVGSFADYAPQNDILMLGAVGAFOWSGTLVQETSHKPIV-----FPK 418
QY 417 EL-----KNHGAYLGYTVTSVVSSRQGRVYVAGAPRENHTGKVLFTMHNRESLTIHQ 469
Db 419 QAFDQVLQDRNHSFLGYSV-AAISTEDGVHVFAGAPRANTYGOVLVYSVKNQGNVTVIQ 477
QY 470 AMRQQTGSGTSGEITSDIDGVDTVLLVGAPMYFNE-GRERGKYVYVELRQNRVYN 528
Db 478 SHRGDQISGYFSGVLCSDVDKDTITDVLVGAPTYMNDLKEEGKVLFTIKGILNQH 537
QY 529 GTLKDSHSYNARFGSSIASVRDLNQSYNDVVGAPLEDNHAGAIYIFHFGRSILKTP 588
Db 538 QFLEGGEGTGNARFGSAIALSDINMDGFDNDVIIVGSEVENENSGAVYIYNGHQTIRTKY 597

```

```

QY 589 KORITASELA--TGLQYFCSTHGOIDINEDGLIDLAVGALGNVILMSRPVQVNASLH 646
Db 598 SOKILGSGAFRRHLQFFERSLDGYDLNGDSITDVSIGALGOVLQWQSQSIADVAIEAL 657
QY 647 PEPKINIFHRDKCSGRDATCLAAFLCFTPIFLAPHFQTTTGTGIRYNATMD-----ERY 702
Db 658 FTDPKITLLNKDAK-----ITKLCFRAEF-RPAGONNOVALFNMTLADAGHSSRV 708
QY 703 TPAHLDEGGDRFTNRAVLSSGQELCEFRINPHVLDTADYVKPVTFFSVEYSLEDPDHGPM 762
Db 709 TSGVFRNSERFLQKNMVVNEVQK-CSEHHISIQKPSDVVNPDLRDVDSLENCTSA 767
QY 763 LDDGWPTTLRV-SVPPFWNGCNEDEHCVDPDLVDARSDLPTAMEYCORVLRKPAQDCSAYT 821
Db 768 L-BAYSETVKVFSIPFYKECGSDGICISDLILDVQ-QLP-----804
QY 822 LSFDTVTIIESTRORVAVENTLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIETC 881
Db 805 -AIQTQSFIVNQNKRLTFSVILKNRGESANTVVLAEFSENLFPAFSMPVD---GTEV 860
QY 882 VNEERRLOQV-QNVSYPPFRAKAVAPRLDSEFSKIFLHHLELELAAGSDSNERDSTK 940
Db 861 TCEVGSSQKSVTCVGYPAKSEQVTFINFDNLQNLQCAAINFOAFSESQE--TNK 918
QY 941 EDNVAPLRFHLKYEADVLFTRSSLSHYEVKLN-----SSLERYDGIGPPFSCIFRIQ-NL 995
Db 919 ADNSVSLTIPLLYDAELHLTRSTNINFEISSDENAPSVIKSVEDIGPKF--IFSLKVT 976
QY 996 GLFPIHGMKMITIPIATRSNRLKRLDPLTDEA-NTSCNIWGNSTERYTPV-----1048
Db 977 GSAPVSMALVTIHLPOVTKENKPLLYLTGTQDQAGDISC-----TABINPLKLPHTAPS 1031
QY 1049 ----BEDLRAPQLNHSNDVVSINCNIR-LVPNQEIHFHLLGNLWLSLKALKYKSMKI 1103
Db 1032 VSFKNENFRHTKELDCRTTSCNITCWKLKOLHMAEYFIVNTRVWNTFPASTFQT--V 1089
QY 1104 MVNAALORQSPHPIFREDDPSQIEFISK-QEDMQVPIWIVGTLGGLLLALLVLA 1162
Db 1090 QLTAASAIIDITHNPQLFVIEENAVTIPLMIMKPTKEABVPTGVIITGSIAGILLLLAMTAG 1149
QY 1163 LRKLGFPSARRRREPGLDP 1182
Db 1150 LMKLGFFK--ROYKMGQNP 1167

RESULT 4
I45914
integrin alpha 2 subunit - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Sep-2003
C;Accession: I45914
R;Kamata, T.; Puzon, W.; Takada, Y.
J. Biol. Chem. 269, 9659-9663, 1994
A;Title: Identification of putative ligand binding sites within the I-domain of integrin
A;Reference number: A54402; MUID:94193647; PMID:7511592
A;Accession: I45914
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1170 <KAM>
A;Cross-references: GB:I25886; NID:G439695; PIDN:AAB59255.1; PID:G439696
F;161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 30.6%; Score 1901.5; DB 2; Length 1170;
Best Local Similarity 36.3%; Pred. No. 1.4e-118;
Matches 437; Conservative 238; Mismatches 437; Indels 91; Gaps 30;

QY 23 FNMDTKPRVIGSRRTAFGYTVQVQHDISGNKWLVVGAPLENGYQKTGDVYKCPV--IH 80
Db 19 YNVGLPKAKIFSGPSSEQGYAVQVQFINPKGNLWLVGSPWSPFKNRMGDVYKCPDLST 78
QY 81 GNCTKLNLCGRVT-LSNVSRKDNMRGLSLATNPKNDSFLACPLWSHCGSGSYTTGMC 139
Db 79 TTCEKLNLTQTSMSNSNVTEMKNMNLGLTLTRNVGTGGFLTCGLPLWAQCGSGYTTGVC 138

```

```
Qy 1186 VLE 1188
Db 1165 TTE 1167

RESULT 5
A33998
integrin alpha-2 chain precursor - human
N;Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 c
C;Species: Homo sapiens (man)
C;Date: 30-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 15-Sep-2003
C;Accession: A33998; B56793; A53117
R;Takada, Y.; Hemler, M.E.
J. Cell Biol. 109, 397-407, 1989
A;Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit (platelet
A;Reference number: A33998; MUID:8930879; PMID:2545729
A;Accession: A33998
A;Molecule type: mRNA
A;Residues: 1-1181 <TAK>
A;Cross-references: GB:X17033; NID:g33906; PIDN:CAA34894.1; PID:g33907
A;Note: the authors translated the codon GAT for residue 802 as Gln, GTC for residue 803
R;Catimel, B.; Parmentier, S.; Leung, L.L.; McGregor, J.L.
Biochem. J. 279, 419-425, 1991
A;Title: Separation of important new platelet glycoproteins (GPIa, GPIc, GPIc*, GPIIa a
A;Reference number: A56793; MUID:92061944; PMID:1953640
A;Accession: B56793
A;Molecule type: protein
A;Residues: 30-43 <CAT>
A;Experimental source: platelet
R;Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Tsung, Y.L.; Gafford, A.
J. Biol. Chem. 269, 463-469, 1994
A;Title: The human alpha-2 integrin gene promoter. Identification of positive and negat
A;Reference number: A53117; MUID:94103255; PMID:8276836
A;Accession: A53117
A;Molecule type: DNA
A;Residues: 1-16, 'y', 18-21 <ZUT>
A;Cross-references: GB:L24121; NID:g400342; PIDN:AAA16619.2; PID:g4583535
A;Note: authors translated the codon GTA for residue 17 as Leu
C;Genetics:
A;Gene: GDB:ITGA2; CD49B
A;Cross-references: GDB:128031; OMIM:192974
A;Map position: 5q11.1-5q11.2
C;Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-1133/Domain: extracellular #status predicted <EXT>
F;172-347/Domain: von Willebrand factor type A repeat homology <VWA2>
F;1134-1154/Domain: transmembrane #status predicted <TM>
F;1155-1181/Domain: intracellular #status predicted <CYT>
F;1105,112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Asn) (covalent
```

```
Qy 140 SRVNSFRFSKTVAPALQRCOTYWDIVIVLDGNSIYVPWVQVHPLNLIKFKVIGRCQI 199
Db 139 SDVSPDFQRTSFAPVQTCFSFIDVVVVCDESNIYPDWAVKRNFLKFKVQGLDIGPTKT 198
Qy 200 QGVVVQVGEDVVFHFLNDYRSVDVVEAASHIQRGCTETRTAFGLIEFARSEAFQ--KG 257
Db 199 QMGLIQVANNPRVFNLTFRSKDEMIAKATSCOTFOYGGDLTNTFKAIQIARDTAYSTAG 258
Qy 258 GRKGAKKVMIVITDGEHSDSPLEKVIQQRSDNVTRYAVAVLGYNNRRGNPFTFMEI 317
Db 259 GRGATKVMVVTDGEHSDGSKLKAVIDQCNKDLNRGLIAGVLYNENALDTKNLKEI 318
Qy 318 KYIASDPDDKHFFNVNTEAALKDIVDALGRIFSLEGNKNETSPGLEMSQTGSRSHVVE 377
Db 319 KAIASIPTEHFNFVNSDEADLLEKAGTIGEQIFSIETGTVQGDNFQEMMSQVGSABYSP 378
Qy 378 DG--VLLGAVGDWNGAVLKETSAGKVIPLRESYLKEFPEEL---KNHGAVLGYVTTSV 432
Db 379 QNNILMLGAVGAYDSGTVVQKTPHGHLI----FSKQAFEOILODRNHSSYLGYSVAS- 432
Qy 433 VSSRGRRVAVAGAPRNFHTGKVLFTMHNRRSLTIHQAMRGQQIGSYFGSEITSVDIDG 492
Db 433 ISTGNSVHFVAGAPRANTYQGVILSVNENGNTVIQSRGQDQIGSYFGSVLCVAVDNKD 492
Qy 493 GVTDLVLLGAPMYFNE--GREGRKVVYELRQNRVFNVTGLKDSHYQNARFGSSIASVRD 551
Db 493 TITDVLVLLGAPMYMDLKEEGRVILFTITKGLNWHQFLEGPNGLNENARFGSAAALSD 552
Qy 552 LNQDSVNDVVGAPLEDNHAGAIYIPHGPRGSILKTPKQRIITASELA--TGLQYFGCSIH 609
Db 553 INMDGFNDVIGSPLENQNSGAVIYNGHEGMRIRLYSQILGSDRASFSSHLOYFGRSLD 612
Qy 610 GQLDLENDGLIDLAVGALGNVILMSRPVQVNASLHPEPSKINIFHRDCKRSGRDATCL 669
Db 613 GYGDLNGDSITDVSUGAFQVQLMSQSADVSVDASFTEPKITLLNKNAE-----I 664
Qy 670 AAFLCFTPIFLAPHFTTTVIGIRYNATMDERRYT----PRAHLDGGRFTNRAVLSSG 725
Db 665 KLKLCFSAKF-RPTNQNNQVAIVYNTITIDEDQSSRVSIRGLFKENNERCLOKTMVSOA 723
Qy 726 QELCERINPHVLTADYVPKVTSEVYSLEDPDHGMDDGWPTTLRV--SVFPWNGCND 784
Db 724 QR-CSBYIIHQPSDIIISPLNLCMISLENPTNPAL-EAYSETVKVFSIIPHKDCGDD 781
Qy 785 EHCVPDLVADARSDLTAMEYCORVLRKPAQDCSAYTSLFTDTVFVFIESTQRQVAVEATL 844
Db 782 GVCISDLVLNVQ-QLPATQQ-----QP-----FIVSNQNKRLTFSVOL 818
Qy 845 ENGENAYSTVLNISOSANLQFASLIQKEDSDGSIQVNEERLQKV-CNVSYPFPRAK 903
Db 819 KKKESAYNTEIVVDFSENLFASWMPVD---GTEVTCQIASSSQKSVTCNVGYPALKSK 875
Qy 904 AKVAFRLDSFSKSIPLHLEIELAAGSDSNRSDTKEDNVAPLRPHLYKAEADVLTFRSS 963
Db 876 QQVTFITFDNFIQNTQNASISFRALESQENMA--DNSVNLKLSLLYDABIHTRST 933
Qy 964 SLGSHVEKL---NSSLERYDGIQPPPSICFRIQ-NLGLPPIHGMKMITIPATRSGR 1018
Db 934 NINFEVSLDGNVSSVHSPEDIGPKF--IFSIVKTTGSPVPSMASVLIHIPQYTKDKNP 991
Qy 1019 LLKLRLDFTDEA-NTSCNIGNSTERYPTV-----BEDLRAPOLNHSNDVVS 1067
Db 992 LMYLTGVHTDQAGDISC-----EAEINPLKIGTSSSVSPKSENFRHKEJENCRATCSN 1046
Qy 1068 INCNIR-LVFNQENFPHLLGNLWLSLKALKYKSKMIMVNAALQROPHSPFIFREEDPSR 1126
Db 1047 IMCWLRDLQVKEGYFLNVSTIWNGTAASTFQI--VQLTAAAEIDTYNPQIVIEENTV 1104
Qy 1127 QIEFEISK-QSDWQVPTIIVGSTLGLLLALLVLAALRLKLGFRSRRRRREPGLDTPK 1185
Db 1105 TIPTIMKPKHEVPTGVIVGSVIAGILLALLVAILWKLGFPRKRYKAKNPNDETDE 1164
```


QY 818 SAYTLSDFTTTFIIESTRQRVAVATLENRGENAYSTVLNTSQSANIQFAS----- 868
Db 783 TMSANGLDT---LVVGGQDFNMSTVLTNRDGEDSYGTQTVVYPSGLSYRKDSASQNPILT 839
QY 869 ---LIQKEDSDGSIIECVNEERLQKQVNVSYSPFRRAKAKVAFELDSEF-SKSIFLHL 923
Db 840 KKPWFVKPAESSSSSE---GHGALKSTTWNHPIFPANSEVTFNVTVDVDSHASFNGKL 896
QY 924 EIELAAGSDSNEROSTKEDNVAFRLPHLYKEADVLFTRSSLSHY-----EVKLNS 975
Db 897 LLKAVAGENNM---SRTHKTKQELPVKYAIYIMVITDSESRVYLVNTASEMSKVQHQ 954
QY 976 LERYDGIGP---PFSCIF---RIQNLGLFPIHGMKMITTIPIATRSNGRLLLKLRDFTD 1028
Db 955 QYQFNGLGORSPLVSVWFVPIQINNVTWMD-HPQVI-----FSQ 993
QY 1029 EANTSCNTWGSTYRPTP---VEEDLRAPOLNHSNDVVSINIRLVPNOET-NPFL 1084
Db 994 NLSSACH-----TEOKSPHSPFRDQLERTPVLCNSVAVCKRIQCDLPSFTQEIFNVTL 1048
QY 1085 LGNL---MLRSLK---ALKVKMKIMVNAALQRFHSPPIFREEDPSPRQIEFEISKQEDW 1138
Db 1049 KGNLSDFTWIKTSHGHLVLSSTELFN-----DSAFALLPGQESYVRSXTETKVPEY 1101
QY 1139 QV---PIWITVSGTLGGLLLALLVLALRKLGFPR 1170
Db 1102 EVHNPVLIVSGSIGGLVLLALITAGLYKLGF 1135

RESULT 9
RWHU1B
N;Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac-1
eukocyte integrin alpha chain; neutrophil adherence receptor alpha chain
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence, revision 31-Dec-1992 #text change 31-Dec-2000
C;Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; I52567
J;Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.
J. Biol. Chem. 263, 12403-12411, 1988
A;Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b)
B.
A;Reference number: A31108; MUID:88315033; PMID:2457584
A;Accession: A31108
A;Molecule type: mRNA
A;Residues: 1-1153 <COR>
A;Cross-references: GB:J03925; NID:g187284; PIDN:AAA59544.1; PID:g307148
A;Note: part of this sequence was confirmed by protein sequencing
R;Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.
J. Cell Biol. 106, 2153-2158, 1988
A;Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mac-1
A;Reference number: A28915; MUID:88257215; PMID:2454931
A;Accession: A28915
A;Molecule type: mRNA
A;Residues: 1-499,501-965, P', 967-1153 <ARN>
A;Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:g186935; PIDN:AAA594
A;Note: the authors translated the codon TAC for residue 1129 as Thr
A;Note: part of this sequence, including the amino end of the mature protein, was confirmed by protein sequencing
R;Shelley, C.S.; Arnaout, M.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991
A;Title: The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression
A;Reference number: A41600; MUID:92073318; PMID:1683702
A;Accession: A41600
A;Molecule type: DNA
A;Residues: 1-9 <SHE>
A;Cross-references: GB:M76724; NID:g180018; PIDN:AAA58410.1; PID:g553215
R;Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Tenen, D.G.
Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988
A;Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion molecule-1
A;Reference number: A94193; MUID:88190151; PMID:2833753
A;Accession: A30892
A;Molecule type: mRNA
A;Residues: 917-1042 <AR2>

A;Cross-references: GB:M18044
R;Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.
Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989
A;Title: cDNA sequence for the alphaM subunit of the human neutrophil adherence receptor
A;Reference number: A32218; MUID:89098893; PMID:2563162
A;Accession: A32218
A;Molecule type: mRNA
A;Residues: 9-1153 <HC>
A;Cross-references: GB:J04145; NID:g189068; PIDN:AAA59903.1; PID:g386975
A;Note: part of this sequence was confirmed by protein sequencing
R;Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.
J. Immunol. 150, 480-490, 1993
A;Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1 chain during evolution.
A;Reference number: A46526; MUID:93123748; PMID:8419480
A;Accession: A46526
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-499,501-1153 <FLE>
A;Cross-references: GB:S52227; NID:g263047; PIDN:AA824821.1; PID:g263049
A;Note: the last three bases of intron 13, CAG, are included in some but not all mature transcripts
R;Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.
Biochim. Biophys. Acta 874, 368-371, 1986
A;Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across species
A;Reference number: A90664; MUID:87076671; PMID:3539202
A;Accession: A26091
A;Molecule type: protein
A;Residues: 17-31 <PIE>
A;Experimental source: granulocytes
R;Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.
Blood 79, 865-870, 1992
A;Title: Characterization of the myeloid-specific CD11b promoter.
A;Reference number: I52567; MUID:92144986; PMID:1346576
A;Accession: I52567
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-9 <RES>
A;Cross-references: GB:M84477; NID:g180184; PIDN:AAA51960.1; PID:g553219
C;Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1
C;Genetics:
A;Gene: GDB:ITGAM; CR3A
A;Cross-references: GDB:120599; OMIM:120980
A;Map position: 16p11.2-16p11.2
A;Note: promoter contains a GATA motif and two Spl consensus binding sites
C;Superfamily: cell surface glycoprotein C1b; von Willebrand factor type A repeat homology
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-1153/Product: cell surface glycoprotein C1b #status experimental <MAT>
F;148-318/Domain: extracellular #status predicted <EXT>
F;148-318/Domain: von Willebrand factor type A repeat homology <VWA2>
F;465-473/Region: calcium/magnesium binding #status predicted
F;530-538/Region: calcium/magnesium binding #status predicted
F;593-601/Region: calcium/magnesium binding #status predicted
F;1109-1134/Domain: transmembrane #status predicted <INT>
F;1135-1153/Domain: intracellular #status predicted <INT>
F;86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Bind
Query Match 17.6%; Score 1094; DB 1; Length 1153;
Best Local Similarity 28.6%; Pred. No. 1.4e-64;
Matches 353; Conservative 208; Mismatches 517; Indels 158; Gaps 46;
QY 12 ALSILWPGFTDTENNDRKPRVIFGSRTPAFFGYTVQQHDISGNKWLWVVGAPLEFNGYQKGT 71
Db 10 ALILCHG----FNLDTENAMTFQENARG-FGQSVQ--LQSSR-VVVGAPQEIIVAAANQRG 61
QY 72 DVYKCPVHGNCTKLNIGRVTLGNVSKRKNMRLGLSLATNPKNDSPLACPLWSHSCGS 131
Db 62 SLVQCDYSTGCEPIRL-----QVPVEAVNMSLGLSLAATSPPLQACGFTVHQTCSE 115
QY 132 SYTTGMCVRVNSNFRSKIVAP-ALQRC-QTYMDIVILVDGNSIYP--WVEVQHFLIN 187
Db 116 NTYVKGICLFLGSLNLRQOPQKFPBALRGCPQEDSDIAFLIDGSGSIIPHDFRMKEFYST 175

686 -TTTGVIRYNATMDERRYTPRAHLEGGDRFTNRAVLSSQELCERINFHV-LDTADYV 743
663 GRILLANLTYTLQDCHRMRSGLFPDGSHELSGNTSI--TPDKGLDFHFFPFCIQDLI 720
744 KPVTFSVEYSLEDPHGMLDDG-----WPT--TLRVSPFWNGCNEDEHCVPLDVLVD 794
721 SPINVSFLNLEEBEGTPRQGRAMQPIILPSIHTVKEIPFEKNCEDKKCANLTLS 780
795 --ARSDLTAMEYCORLVRKAQCSAYTLSDFTTTFIESTROFVAVAEATLENGENAY 852
781 SPARKS-----PURLMSSASL-----AVEWILNSGDAY 810
853 STVLNISQANLQF--ASLIQ-----KEDSGSIECVNEERLQKV--CNVSYPPF 900
811 WURLDLPRLGSPRKVEMLOPHSRMPVSCBELTEGS-----SLLTKTLCNVSSP 863
901 RAKAKVAFR-----LDSEFSKIFLH--HLEIELAAGSDNSNERDSTKEDNVAFLRPHL 951
864 KAGQVSLQVMENTLLNSWEDFVELNGTVHCE-----NENSSLOEDNSAATHIPV 914
952 KYEADVL-----FT-----RSSLSH-YEVKLNSS-----LERYDGIQPPF 986
915 LYPNVLKKEOENSTLYISFTPKGPKTQVQVHVQVRIOPSAYDHNMPTLEALVGVPRPH 974
987 SCIFRQNLGLPPIHGMKMTIPIATRSNRLKLRLDELTEANTSCNINWGNSTEYRPT 1046
975 S-----EDLIT-----YTVSQVTDPLVT 992
1047 PVEEDLRAPQLNHSNDVVSINCINRLVNPNOEINFHLLGNLWL-RSLKALKYSK--KI 1103
993 CHSEDLKR-PSSEAEQCLPFGVOFCPIVFRWEILLQVGTVELSKEIKASSTLSLCSL 1051
1104 MYNAALQROFSPFIEEDPSPQIEFEISKQEDMQVPIIIVGSLGGLLALLLVLAL 1163
1052 SVSFSNKKIFH---LYGSKASERQVLKVDLIEKEM-LHYVLSIGIGLVLFLFLAL 1107
1164 RKLGF-RRARRRE 1177
1108 YKVGFFKRLKEM 1122

RESULT 11
A53213
integrin alpha-E chain - human
C:Species: Homo sapiens (man)
C>Date: 19-Oct-1995 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: A53213
R:Shaw, S.K.; Cepek, K.L.; Murphy, E.A.; Russell, G.J.; Brenner, M.B.; Parker, C.M.
J. Biol. Chem. 269, 6016-6025, 1994
A:Title: Molecular cloning of the human mucosal lymphocyte integrin alpha (E) subunit. Un
A:Reference number: A53213; MUID:94164962; PMID:8119947
A:Accession: A53213
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1179 <SHA>
A:Cross-references: GB:L25851; NID:9457244; PID:9457245
C:Genetics:
A:Gene: ITGAE
A:Cross-references: GDB:330801
A:Map position: 17p13
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom
F:199-371/Domain: von Willebrand factor type A repeat homology <WAA>

Query Match 14.9%; Score 924.5; DB 2; Length 1179;
Best Local Similarity 26.3%; Pred. No. 3.2e-53;
Matches 338; Conservative 206; Mismatches 482; Indels 259; Gaps 52;
23 FNMTRPRVTPGRTAFPGYVQVQHDISGNK-WLVVGAPLETNGYKGTGVYKCPVHG 81
19 FNVDVAPWLPFGKGAFLVSLSLHQDDESTQTLVLTSP---RTKTPGLHRCSLVOD 75
82 NCTKLNGRVLTSNVSEKDNMRNLGLSLATNPKNISFLACPLW---SHECGSSYYTTGM 138

76 EILCHPVEHVPIPKGRHR-----GVTWRS--HHGVLIICIVLVRPHSLSSB--LTGT 125
139 CS-----RVNSNPFESK-----TVAPALQRC----- 159
126 CSLGLPDLRPOAQANFFDLNLDLPDARVDTGDCYSNKEGGEDDVNTARQRRALEKEEE 185
160 -----QTYMDIVIVLDGNSIYP--WVEVQHFLINILKKFYIGBQIQGVVVQ 205
186 EDKEEEDDEEEEAETAILIDGSGSIDPDPFORAKDFISNMNRNFYKCFECFNALVQ 245
206 YGEDVHFEHLNDVRSVKVVEAAASHIEQRGGTETRTAFGIEPARSEAF--QKGGKCAK 263
246 YGGVIQTEFDLRDSQDVMAASLARVQNTQVGSV-TKTASAMQHVLDLSITSSHSGSRKAS 304
264 KMLVITDGSHTSP-DLEKVIQOSERNVTRYAVAVLGYNNRRGINPETF-----LNEI 317
305 KMWVLTDDGGTFEDPLNLTIVINSFKMGQVERFALGV-----GEEFKSARTAREL 354
318 KYIASDPDDKHFFNVTDDEAALKDIVDALGDRIFSLSEGTNKNETSFGLEMSQTFGSSHVE 377
355 NLIASDPDETHAFKVTNYMALDGLLSKLYNIISMEGTVGDAHY--QLAQIGFSAQILD 412
378 D-GVLLGAVGAYDNW-GAVLKETSAGKVIPIRESYLVKEFPPELKNHGAVLGYTVTSVSS 435
413 ERQVLLGAVGAFDMSGGALLYDTRRRGRFLNQTAADADAQAQY-SVLGYAV-AVLHK 470
436 RQGRVYVAGAPRFNHTGKVFLLFTMNNRSLTIHQAMRGOQIGSYFGEISITSDIDGVT 495
471 TCSLSYVAGAPQYKHGAVPEL-OKEGREASFLPVLGEQMGSYFSGELCPVIDMDGST 529
496 DVLLVGAPMYFNEGRERKGVVYVEL--RQNRVYNGTILKDSHYONAPRGSSIASVRDLN 553
530 DFLVAAAPYHVH-EEGRVYVYRLSEQDGSFSLARILSGHPFTNARFGFAMAAGDLS 588
554 QDSYNDVVVGAPLE---DNHA--GAIYIFHGRSILKTPKORITASLQATLGLOVFGCS 607
589 QDKLTVAITGAPLEGGADGASFGSVIYNGHWDLGSLASPSQRIRASTVAPGLQIFGMS 648
608 IHGQLDNEEDGLDLAVGALGNVILWSRPVQVINAHLFEFSPKINIFHRDCKRSGRDAT 667
649 MAGGFDISGDLADITVGLGQAVFRSPVRLKVSMAFTPSALPIFGNG----- 699
668 CLAAFLCFTPIFLAPHFQTTTVGIR---YNATWDERRYTPRAHLD-----E 710
700 VVNVRLCFB--ISSVTTASEGLREALNFTLDVDVGKQRRRLQCSVRSCLGLCREWS 756
711 GGDRTFNRAVLSSGGELCERINFHVLDTADYVKPVTFSVEYSLEDP---DH-GPMLDD 765
757 SGSQLCEDILLMPETEGELCEE-----DCFSNASVKVSYQLOQTEPGQTDHPQPLDR 807
766 GWFTTLRVSPFWNGCNEDEHCVPLDVLDAKSDLTAMEYCORVLRKPAQDCSAYTLSD 825
808 YTEPFAIFQLPYEKACKNKLFCVAEIQL-----ATTVSQQ 842
826 TTVFIIESTRQVAVAEATLENGENAYSTVLNISQANLQFASLIQKEDSGSIECVNEE 885
843 E---LVVGLTKELTLNLTNSGDSYMTSMALNPRNLQKRM-QKPPSP-NIQCDPQ 897
886 RRLQKV--CNVSYPPFR---AKAKVAFRLDSEFSKIFLHLEIELAAGSDNSNERDSTK 940
898 PVASVLIMNCRIGHPRVKRSSAHVSVVWQLE---ENAFNRTADITVTVNSNRRSLA 953
941 EDNVAPLRFHLKVEA-----DVLFTRS--SSLSHYEVKLNSSLERYDGIQGFPCIFRQ 994
954 NE-THTLQFRHGFVAVLSKFSIMVYNTGQGLSHHKE----- 988
995 LGLPPIHG-----MMKTIPIATRSNRLKLRLDFTDEANTSCNIWG--NSTEYRP 1045
989 -FLFHVHGENLFGAEYQIQCVPTKLR-GLQVAAVKKLTRTQASTVCT-NSQERACAYSS 1045
1046 TPVEEDLRAPQLNHSNDVVSINCINRLVNPNOEINFHLLGNLWL-RSLKALKYKYS 1100
1046 VQHVVEHWSVCVTASDKENVTVAAEISWDHSEELLKDVTELQILGEISFN-----KS 1098

QY 1101 MKINVAALQORHSPPIFEEDPSRQIEFEISKQEDWQVPIWIIIVGSTLGLLALLV 1160
 Db 1099 LYEGLNA-----ENHRTKITVVLKDEKYH-SLPILIKSGVGLLVILVIL 1143
 QY 1161 LALRKLGFERSARRRREPGLDTPK 1185
 Db 1144 VILFKCGFFK--RKYQOLNLESIRK 1166

RESULT 12
 I58409
 Integrin alpha-9 chain precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 31-Dec-2000
 C:Accession: I58409; A49459
 R:Hihi, K.; Yamakawa, K.; Ueda, R.; Horio, Y.; Murata, Y.; Tamari, M.; Uchida, K.; Takah
 Oncogene 9, 611-619, 1994
 A:Title: Aberrant upregulation of a novel integrin alpha subunit gene at 3p21.3 in small
 A:Reference number: I58409; MUID:94119603; PMID:8290272
 A:Accession: I58409
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-1035 <RES>
 A:Cross-references: GB:D25303; NID:9464180; PIDN:BAA04984.1; PID:g533327
 R:Palmer, E.L.; Ruegg, C.; Ferrando, R.; Pytela, R.; Sheppard, D.
 J. Cell Biol. 123, 1289-1297, 1993
 A:Title: Sequence and tissue distribution of the integrin alpha9 subunit, a novel partne
 A:Reference number: A49459; MUID:94064789; PMID:8245132
 A:Accession: A49459
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 30-1035 <PAL>
 C:Superfamily: integrin alpha-4 chain
 C:Keywords: glycoprotein; metal binding, transmembrane protein
 F,1-27/Domain: signal sequence #status predicted <SIG>

Query Match 12.8%; Score 799.5; DB 2; Length 1035;
 Best Local Similarity 24.2%; Pred. No. 6e-45;
 Matches 307; Conservative 177; Mismatches 418; Indels 369; Gaps 49;

QY 5 RGLVVAWALSWPGETDTFNDTRKPRVIPSRTAFFGYTQQHDSIGNKWLIVGAPLET 64
 Db 14 RALLIALVAVAGIP--AGAYNLDLPORVHFQPADSFFCYAVLEPHDNTRWLVGAPKAD 71
 QY 65 NGY-----OKTGDVYKCPVHGN---CTKMLGRVTLNVNS-----ERKNMRLGLSLA 110
 Db 72 SKYSPSVKSPGAVFKCRV-HINPDRCTELDMARKNGRTSCGKTCTREDRDDWGVSLA 130
 QY 111 TNPK-DNSFLACSPLWSHECGSSYYTT-----GMCGRVNSNFRF-SKTAVAPLQRCQTY 162
 Db 131 RQPKADGRVLACAHWK---NIYEADHILPHGFCVLIIPSNLQAKGRTLIPCYEE--- 182
 QY 163 MDIVIVLDGNSIYPWVEVQHFNLINLKKFVIGQIQGVGVQYGEDVWHEFHLDYRSV 222
 Db 183 -----YKKKY----- 187
 QY 223 KDVVEAASHIEQRCGTETRTAFGTEFARSEAFQKGRGAKVMIVITDGEHSDSPLEK 282
 Db 188 -----GEEHGS----- 193
 QY 283 VIQOERDNVTRYAVAVLGYNNRRGINPETPLNBIKYIASDPDKHFFNVTDAAALKDIV 342
 Db 194 ----- 193
 QY 343 DALGDRIFSLEGTWKNETSFGLEMSQTFSSHVVEDGVLLGAVGAYDNGAVLKETSAGK 402
 Db 194 -----CQAGIAGFFTEELVNVGAPGSPFWAGTI-----K 222
 QY 403 VIPLRE-SYLKERPEELKNHG-AVLGYTVTSVSVSSRGRV-YVAGAPRFNHTGKVLFTW 459
 Db 223 VLNLDTNTYKLNDVIMNRRYTYLGVAVTAGHFHSPHTIDVVGGAQDKGIGKVIFRA 282

QY 460 HNNRS---LTIHQMRGQQIGSYFGSEITSDIDGQVTDVLLVGVAPMYNEGRERGVY 516
 Db 283 -DRSSTGLIKIFOA-SGKMGSYFGSSLCVAVDLNGDGLSD-LLVGAPM-FSEIRDEQV 338
 QY 517 VYELRONRFVYNGTLKDSHYQNAFGSSSTASVRDLNQDSYNDVYVGVAPLEDNHAGAIYI 576
 Db 339 VYINRNGALAEQLALTGCAYNAHFGESIADLDDNDGFPDVAIGAPKEDDFAGAVYI 398
 QY 577 FHGFRGSILKTPKQRIITASELATGLQYFGSIHGOLDLNEGLDIDLAVGAL--GNAVILW 634
 Db 399 YHGDAGGIVPQYSMKLSGOKINPVLPMFGQSIHGGIDMGNGVDPDVTVGFAMSDSVLLR 458
 QY 635 SRPVQVINALSHPEPSKINIFHRDCKRSRGRDATCLAAFLCFTPIFLAPHFQITTVGIRYN 694
 Db 459 ARPVITVDVSI-FLPGSINITAPQCHDGOQPVNCLNVTTFCF--PHGKHV-PEEIGLVY 514
 QY 695 ATMD---ERRYTPTAH---LDEGGRFTNRAVLSSGOELCERINFHV-LDTADYVKPV 746
 Db 515 LMADVAKKEGOMPRIYFVLLGETMGQVTEK-LQTYMEETCRHYVAHVKKRVQDVISPI 573
 QY 747 TFSVEYSLED-----PDHGPM--DCGPTTLRVSVPFMNGCNEDEHCVDPDLVD 794
 Db 574 VFEAAYSLSSEHYTGEEERELPLTLPVLRWKKGOKIAQKNQTVFERNC-RSEDCAADLQ 632
 QY 795 ARSDLTAMEYQORVLRKPAQDCSAYTLSFDTTVFIESTRQORVAVEATLENRGENAYST 854
 Db 633 GKL-LLSMD-----EKTLYLALGAVKNISLINSISNLGDDAYDA 671
 QY 855 VNISQSANLOFASLLOKEDSDGSIETCVNEERLQOVNVSYPFFRAKAKVAFRLDSEF 914
 Db 672 NVSNVNSRELFFINWQKEE--GISCELLESDFLK--CSVGFPFMRSKSKY-----EF 721
 QY 915 SKSIFLHLEIE-----LAAGSDSNDRDSTKEDNVAFLRPHLYKVEDVLT--RSSSL 965
 Db 722 SVIPDTSHLSGEEVLSFIVTAQSGNTERSESLHNTILVMVPLMHEVDTSITGMSPTS 781
 QY 966 SHYEVKLNS---LERYDGGPPFPFCIFRIONLGLFFIHGMMKITIPIATRSN-RLL 1020
 Db 782 FVYGESVDANFIQLDDLECHFPQINITLQVNTGPTSLPGSSVSISFNLSSGAEWF 841
 QY 1021 KLRFDLTDEANTSNIGNSTEYRTPV-----BEDL-----RRAPQLNHSNS 1063
 Db 842 HVQEVVQEGKNGCSFQKN-----PTPCIIPQEQENIFHTIFAFFTKSGRKVLDCEKPGI 896
 QY 1064 DVWSINCINIRLVNQE---INFHLLGNLWLSRLKALKYKSKMIMVNAALQORHSPPIFR 1120
 Db 897 SCLTAHNFSALEESERTIDIYMLN-----TEILKXDSVVI-----QFWSRAKVK 944
 QY 1121 BEDPSRQIEFEISKQEDWQVPI-----WIVGSTLGLLALLVALLRKL 1166
 Db 945 VDPALRVVETAHGNPEEVTVVFEALHNLPRGVVGVWIIAISLLVGLIFILLAVLLWKM 1004
 QY 1167 GFFRSARRRRE 1177
 Db 1005 GFPR--RRYKE 1013

RESULT 13

A55348

integrin alpha-1 - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 15-Sep-2003

C:Accession: A55348

R:Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.

J. Biol. Chem. 269, 22811-22816, 1994

A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1

A:Reference number: A55348; MUID:94357930; PMID:7521332

A:Accession: A55348

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <KER>

A:Cross-references: GB:U0114

Tue Sep 21 14:48:56 2004

us-09-980-403-2.rpr

F;55-230/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 12.7%; Score 789; DB 2; Length 272;
Best Local Similarity 55.9%; Pred. No. 3.2e-45;
Matches 151; Conservative 53; Mismatches 62; Indels 4; Gaps 3;

QY 108 SIATPKNSFIACPLMSHEGSSYYTGMCSRNSNFRSKTVAPALQRCOTYMDIVI 167
DB 2 TLVITNPK-GGFLACGLVAYKGRLLHYTGVCSNVSSTFETVAVAPVQCEKTLQDIVI 60
QY 168 VLDGNSIYPPVEVOHFLINILKXYIGQIQGVQVOYGEDVHVEFHNDYRSVKVVE 227
DB 61 VLDGNSIYPPVESVAFNLNSLRNMDIGQQQGVQVIGVQGVVHVEFNLNYSYTFEEVMD 120
QY 228 RASHTEQGGTETRTAFGIEFARSEAFQK--GCRKGAKKVMIVITDGESHSDPDLEKVIQ 285
DB 121 AALRTQGGTQMTALGIDTAREEAFTEAHGARRGVQVMVIVITDGESHSDNYRLQEVID 180
QY 286 QSERDNTYAVAVLGYNNRRGINPETFLNEIKYIASDDDKHFFNVTDAAALKDIVDAL 345
DB 181 KCEDENIQFAIALIGSYSRGNLSTKPFVEIKSIASKPTKHFNVSDALALVTIVEAL 240
QY 346 GRIEISLEG-TNKNETSFGLEMSOTGFSSH 374
DB 241 GERIFALEATTQQAASFEMEQAGFSAH 270

RESULT 14

T31437
Integrin alpha chain SU2 - sea urchin (Lytechinus variegatus)
C;Species: Lytechinus variegatus (variegated urchin)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T31437
R;Hertzler, P.L.; McClay, D.R.
Submitted to the EMBL Data Library, May 1998
A;Description: Alpha SU2, a sea urchin integrin which binds laminin.
A;Reference number: Z21035
A;Status: preliminary; translated from GH/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1041 <HER>
A;Cross-references: EMBL:AF067658; NID:g3220240; PID:g3220241; PIDN:AAC23572.1
A;Experimental source: developmental stage embryo
C;Function:
C;Description: binds laminin
C;Superfamily: integrin alpha-2b chain

Query Match 11.9%; Score 738; DB 2; Length 1041;
Best Local Similarity 23.7%; Pred. No. 7.8e-41;
Matches 308; Conservative 168; Mismatches 422; Indels 402; Gaps 51;

QY 3 LPRGLVAVALSMPGFTDFNMDTRKPRVIGPSRTAFFGYTVQOHDISGNKMLVWGP- 61
DB 1 MARILLISFYAILLDSTAGFNFLRAPLKFDPGDSMGFSPVAQRDGNITDVALVGAPE 60
QY 62 LETN--GYQGTGDYK-CPVIHGN--CTKMLGRV-TLSNVSEKDNRLGLSLATPKD 115
DB 61 ASTNOSGVTGGAVYSCVRLSPGSPVCEQIPFDRTGINNVYD--KSNQWFGATVAGGAN 118
QY 116 NSFLACSP--LWSHECGSSYYTTCGCSRNSNFRSKTVAPALQRCOTYMDIVIVLDGSN 173
DB 119 GQILACAPRLVW----- 130
QY 174 SIYPWVEVQHEFLINILKXYIGQIQGVVQYGEDVVHVEFHNDYRSVKDVVRAASHIE 233
DB 131 -----VETASVTM 138
QY 234 QRGGTETRTAFGIEFARSEAFQKGRKGAKVMIVITDGESHSDPDLEKVIQSERDNTV 293
DB 139 RQEREPTGTCTF----- 149
QY 294 RYAVAVLGYNNRRGINPETFLNEIKYIASDDDKHFFNVTDAAALKDIVDALGDRIFSL 353

150 -----VGSDFTNFVNYS-----PCQ 165
QY 354 GTNKNETSFG-LEMSOTGFSSHVVED--GVLLGAVGAYDMNGAVLKETSACKVPIPLRESY 410
DB 166 STNEDYGFDKITHCQAGFGIIFSDNSALVWGPAGSYLQGIYVQSL-----LNRSV 219
QY 411 LKPEPEELKNHG-----AYLGYYTVTSVSSRQG--RVYVAGAPFNH--TGKVVILTMNN 462
DB 220 VQATQE--SNTGTYSFONSIEGYSIALGDFNGDGVQDVVGTVPRAESLMGLVAIPDQNLN 277
QY 463 RSLTIHOMRQOQGSYFGSITSVDIDGDGTVTVLLVNGAPMYFNEGR-----ERGVVV 517
DB 278 Q---FNQVM-CTQIVAYFGSYTVVDINNDYDD--LLVGAPMYMDGGAIRORWEAGAVIV 332
QY 518 YELR-----QNRVYNGTLKDSHYQNARFGSSIASVROLNQSDYNDVVVVGAPLEDNH 570
DB 333 YLQNDPVGPGASNRLSLSTLIGQI--RSRFGLSIASIGSDNQDGFNDVAIGAPVEGDD 390
QY 571 AGAIYIFHGFRGSIILKTPKORITASELA-TGLOYFGCSIHGOLDLNEGLDILAVGA--L 627
DB 391 AGAVIYHGSANGLKSTPAQVLTPTSLGHSGITTFGLQGGQDMKPKYDPLLVAESA 450
QY 628 GNAVILMRPVPVQVNASLHFESKINIHRDCKRSGRDATCLAAFLCTFPIFLAHPFQT 687
DB 451 NTAVLIRTPVVSLDATLNTPEIGINLENKTYELA--DGTWVTSFIAMTCFTYTGNYLPD 508
QY 688 TVGIRYNATMDERRVTPR--AHLDEGGDRFTNRVALLSSGQELCERINFHVLDT-ADYVK 744
DB 509 HIDISYTVTVDSGIITANRRAMFVNDMSEITKTRRLAVSTQ-FCDFLRAYVGNSEDKLT 567
QY 745 PVTFSVEYSLEDDPHG-----PMLDDGWPPTLRVSPVFWNGCDEHDEICVDPLVDARS 797
DB 568 PIKVTQYLDLNDESLQHEILFIIDMATMSTQTKQVSIQNNC-VNNTCIPDLVDVTVTP 626
QY 798 DLPAMEYQCVLRKPAQDCSATLSFDTVTIIESTRORVAVEATLENRGENAYSTVLN 857
DB 627 NLPN-----IVIGQBELTLDVSLNNGEDAFQSSLS 658
QY 858 ISQSANLQFASLIQKSDSGSIECVNEERRLOQVCNVSYPPFPRAKAKVAFRLD-SEFSK 916
DB 659 VYPLGLQFVRLERKANMDFSVTC-SEDSDLRIITCDTGNPMV-GKNILEFGLTLSTFQV 716
QY 917 SIFLHLEIELAAGSDNSRDSKEDNVAPLRPHLKYEADVLFTRSSLSHYEVKLNSSL 976
DB 717 SGDKDSIEFYFKAENSEDPNTLENNELNMTVPVTVDCITL---KLISASYPRIVMYSTQ 773
QY 977 ERYDGGIPPFSC-----LFRIONLGLPFIHGMKMITIPIATRSNRLKL 1022
DB 774 EDY--VVPFPPAKNASEADIGMBVMHLYEVRNTGSSNAGEVSLNIQWPKNEDGEYLFYL 831
QY 1023 RDLFTDEANTSCNI-----WGNSTYRTPVVEEDLRAPOLNH--- 1060
DB 832 LGIMTEGVT-CQLTQKANKPEGVKLEPSTKAKLSNST---TQVSGRKRREPEVAELA 886
QY 1061 -----SNSDVVSNINCINRLVNPQBIN-----PHLLGNLWLSLK----- 1094
DB 887 QTDNVIYCASDSCVLNCTI-----DEINASKKVVIRILGRFWERTFQKAVSELTVPVQA 941
QY 1095 -----ALKYKSMKIMVNAALORQH-----SPEIFREE--DPSQIEFISQEDWQ 1139
DB 942 TIASSASAAVK---TIPYNIPLRDPDSSTKASTLTVELVDPVTFIA-----W- 988
QY 1140 VPIWIVGSLTGLGILLALLVLRKLGFRRSARRRREP 1179
DB 989 ---WIVVSVLGGIILLIILGLWKCGFF---ERKPKG 1021

RESULT 15

JC7294
alpha integrin - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C;Accession: JC7294

R;Susan, J.M.; Just, M.L.; Lennarz, W.J.
Biochem. Biophys. Res. Commun. 272, 929-935, 2000
A;Title: Cloning and characterization of alphap integrin in embryos of the sea urchin Sc
A;Reference number: JC7294
A;Contents: Embryo
A;Accession: JC7294
A;Molecule type: mRNA
A;Residues: 1-1054 <SUS>
A;Cross-references: GB:AA55724
C;Genetics:
A;Gene: snalphap
C;Superfamily: integrin alpha-2b chain
C;Keywords: calcium binding; embryo; glycolysis; glycoprotein; heterodimer; transmembran

Query Match 11.6%; Score 719.5; DB 2; Length 1054;
Best Local Similarity 23.1%; Pred. No. 1.4e-39;
Matches 292; Conservative 178; Mismatches 428; Indels 365; Gaps 48;

QY 20 TDTNMDTKPRVPGSRFAFGYVQQHDSGNKWLWVGAPLE-----TNGYQKTDG 72
Db 19 TVAFNFDLRAPVKFDGPGQSLFGFSAQHRDQNTDNLVIGAPEAPTQFGVING---GA 74

QY 73 VYKCPVI---HGNCCKNLGRVTLNVSERKDNKRLGLSLATNPKDNFLACSP--LWS 126
Db 75 VYKCPVTPLSGSGPCQCEQVDFDTGTGTEVLVD--KSNQWFGATLASSGPDGRIACAPRLVWL 133

QY 127 HECSSSYTTGMCVRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPPWEVQHFLI 186
Db 134 Q-----TSISPT----- 141

QY 187 NILKFEYIGPGQIQGVVQVQYGEDVHVEPHLNDYRSVKDVVEAASHIEQRGGTETRTAFGI 246
Db 142 -----DKREPTGTCF-- 152

QY 247 BFARGEAFQKGRKGAKKVMIVITDGHSDSPDLKVIQOQSERDNVTRYAVAVLGYNNR 306
Db 153 ----- 152

QY 307 GINPETFLNEIKYIASDDDKHFNVTDEAALKOIVDALGDRI FSLGNTKNKNETSFGLEM 366
Db 153 -VGHSDFTNFVNYSPCQSTDRDLF-----GFDK----- 182

QY 367 SQTGFSSHVVDG--VLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEFPPEE-LKNHGA 423
Db 183 CQAGFSAQIPSDNSTLVMGAPGYSYVLOQIQIFAQSL-----TLSDVSNTPQAVAFDNS 236

QY 424 YLGKVTVTSVSSRQ--RVYVAGAPFNH--TGKVLFTMHNRSLT--IQAMRGQOIGSYF 480
Db 237 YRGYSLALGDNGDGLDYVVGTPRGESLRLGLVAIF-----DQSLVEIITPVVGEQIVSYF 292

QY 481 GSEITSVDIDGVTVDVLLVGAPMYFN-----EGRERKGVVYVELRQNRVYNGTLKDSH 535
Db 293 GYSVASVDVNGDGLDD--LLVGAPMTNREPATEKEAGRVVY--LQNA-----DH 340

QY 536 SYQ-----NARFGSSIASVRDNLQDSYNDVNVVVGAPLE--DNHAGAIYIFHGRGSI 584
Db 341 SLGAPQMLTGKKIRARFGFPITISGDSNQDGFNDVAIGAPYDGEDNSGVVYIYHSAEGL 400

QY 585 LKTPKQRTASELA--TGLQYFGCSIHGOLDNLNEDGLIDLAVGA--LGNVILMSRPVQOI 641
Db 401 RLTESQVLTPELGFSDITTFGFSVDGQDMQDNDYDPLVVGARSADAAILVTRPVVLL 460

QY 642 NASLHFEPSKINIFHRDCK--RSGRDATCLAAFLCFTPLFLAPHFTTTVGRYNNATMDER 700
Db 461 EAELTIBPIGINLKNKIYELPDGVTWTSFVAMACFT--YTG NHL--PARIGLSYTLTVDS 517

QY 701 RYTPRAHL---DEGDRFTNRAVLLSSGQELCERINFHVLDT--ADYVKPVTFSVEYSLED 756
Db 518 ITSGRRALLEVNELSVQTKNRN--LDVGMKPCDPLRAVVVNTIQDKLTPITAVDLQYELTD 575

QY 757 P-----DHGPMLDGWPHTTLRVSPFPWNGNEHCVPDLVLDARSDLPTAMEYCORVL 810
Db 576 ESILLPFIPLINKEAVSSQTKQVSIQNNC--VNNICIPEIGITVTPNLN----- 625

QY 811 RKPQDCSAYTSLFDTVTIIBSTRQVAVEATLENNGENAYSTVLNISQSANLQFASLI 870
Db 626 -----IVIGQAQELTLVVSINNKGEDAFQSTLAVVYPEGLOYVRLE 666

QY 871 OKEDSDGSIETCVNEERLQKQVNSYPPFRKAKVAFRLD--SEFSKSIPLHLELELAA 929
Db 667 RRANMDFSVTC--TEDSALRMITCDTGNPLV--GKYNLEFGLTTLTQVSGDKDNIEFVLA 724

QY 930 GSDSNERDSTKEDNVAPIRFLKYEADVLFTRSS--SLSHYEVKLNSSLERYD----- 980
Db 725 GSENNEDPNTLNNELNVTVAVIVDAILKLLSASYPEIYTVRPEDNVIPEFTKNASEA 784

QY 981 GIGPPFSCIFRIQNLGLFPIHGMKMITIPIATRSNRLKLRLDFLTDEANTSCNIWGS 1040
Db 785 DIGMEVHLYEVRTGSSNAAEVTLNIRWPEKDNGDYLFYLLGIMTDSGVT--CQI--SQ 841

QY 1041 TEYRTPVEEDLRRAPQLNHSNDV-----VSINCN1- 1072
Db 842 GOANPLGVKLEASTKEQJNSNTQVSGRRKREGEVAEALAAQAEPIFCTPESCVLINCTID 901

QY 1073 RLVPNQEIINFHLLGNLWLRSL--KALKYKSMKIMVNA--LQRPQHSPIFEREE 1122
Db 902 EIKATKSKVVRILGRFWERTFOKAVSEAVPVIOVTLASTATATVRSIPYNIPLPWEFTDS 961

QY 1123 DPSRQIEFEISKQEDWQVPI-----WIVGSTLGGILLALLVLALEKLGFFESARRRREP 1178
Db 962 TKASTL---ITABELVLFPVSIAMWIIVSVLGGIILLIIILGLWKCGFF---ERKKP 1014

QY 1179 GLD 1181
Db 1015 GED 1017

Search completed: September 21, 2004, 13:05:00
Job time : 67.7318 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:51:12 ; Search time 179.196 Seconds
(without alignments)
2091.769 Million cell updates/sec

Title: US-09-980-403-2
Perfect score: 6224
Sequence: 1 MDLPRGLVAVALSMPGFT.....FRSARRRBPGLDPTPKVLE 1188

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvrius.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	5690	91.4	1188	11 Q7TQC3	Q7TQC3 mus musculus
2	4272	68.6	823	4 Q8WY18	Q8WY18 homo sapien
3	3840	61.7	823	11 Q8CB84	Q8CB84 mus musculus
4	2135.5	34.3	1171	13 Q42094	Q42094 gallus gall
5	1196	19.2	1161	11 Q9QVE7	Q9QVE7 rattus norv
6	1160	18.6	1169	11 Q9QXH4	Q9QXH4 mus musculus
7	1131.5	18.2	1169	4 Q8IVA6	Q8IVA6 homo sapien
8	1125.5	18.1	1196	13 Q98TF1	Q98TF1 cyprinus ca
9	1096.5	17.6	1151	11 Q9J130	Q9J130 rattus norv
10	1065.5	17.1	1161	11 Q9WTV4	Q9WTV4 mus musculus
11	1063	17.1	1160	11 Q9R200	Q9R200 mus musculus
12	1045	16.8	1187	13 Q98TF0	Q98TF0 cyprinus ca
13	1040	16.7	780	13 Q06271	Q06271 xenopus lae
14	1018.5	16.4	1086	4 Q96HB1	Q96HB1 homo sapien
15	979	15.7	1167	11 Q88341	Q88341 rattus norv
16	954	15.3	1167	11 Q88340	Q88340 rattus norv

17	904.5	14.5	927	6	Q8HZV0	Q8HZV0 bos taurus
18	900	14.5	1160	6	Q8MKF4	Q8MKF4 felis silve
19	895	14.4	920	6	Q28984	Q28984 sus scrofa
20	891	14.3	1038	11	Q8BS01	Q8BS01 mus musculu
21	881	14.2	1332	5	Q9BPQ8	Q9BPQ8 halocynthia
22	856	13.8	835	11	Q9WUF8	Q9WUF8 mus sp. itg
23	848.5	13.6	1036	11	Q8CA73	Q8CA73 mus musculu
24	795.5	12.8	1036	11	Q91YD5	Q91YD5 mus musculu
25	742	11.9	1041	5	Q9UB90	Q9UB90 lytechinus
26	738	11.9	1041	5	Q76378	Q76378 lytechinus
27	719.5	11.6	1054	5	Q9UGS1	Q9UGS1 strongyloce
28	691.5	11.1	1032	11	Q61989	Q61989 mus musculu
29	691.5	11.1	1032	6	Q9BGU3	Q9BGU3 bos taurus
30	667.5	10.7	323	11	Q8CB84	Q8CB84 mus musculu
31	630	10.1	1119	5	Q86G88	Q86G88 pseudopiusi
32	604.5	9.7	1034	13	Q98TT7	Q98TT7 gallus gall
33	597	9.6	632	4	Q8NGH6	Q8NGH6 homo sapien
34	595.5	9.6	1053	11	Q8OYP5	Q8OYP5 mus musculu
35	587	9.4	1132	11	Q8OZ18	Q8OZ18 mus musculu
36	584	9.4	1033	13	Q42598	Q42598 xenopus lae
37	582.5	9.4	1073	11	Q8CC06	Q8CC06 mus musculu
38	571.5	9.2	1474	5	Q86G87	Q86G87 pseudopiusi
39	567.5	9.1	1016	13	Q31779	Q31779 xenopus lae
40	556	8.9	604	11	Q8BQ25	Q8BQ25 mus musculu
41	551.5	8.9	974	11	Q924W2	Q924W2 rattus norv
42	547	8.8	1049	5	Q8SY51	Q8SY51 drosophila
43	545	8.8	1047	6	Q9MZD6	Q9MZD6 bos taurus
44	544.5	8.7	1007	6	Q9GK48	Q9GK48 bos taurus
45	539.5	8.7	833	5	Q9BPQ7	Q9BPQ7 halocynthia

ALIGNMENTS

RESULT 1

Q7TQC3 PRELIMINARY; PRT; 1188 AA.

AC Q7TQC3; DT 01-OCT-2003 (TREMBLrel. 25, Created)

DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)

DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE All integrin.

GN ITGall.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI TaxID=10090;

RN [1] SEQUENCE FROM N.A.

RP Tiger C.-F., Popova S.N., Velling T., Ekman S., Forsberg E.,

RA Gullberg D.;

RA "allbi integrin is important for mesenchymal cell function:

RT elimination of allbi leads to dwarfism";

RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.

RN [2] SEQUENCE FROM N.A.

RP Johansson M., Popova S.N.;

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY124460; AAM62130.1; -

KW Integrin.

SQ SEQUENCE 1188 AA; 133067 MW; 6952BF20D75E4271 CRC64;

Query Match 91.4%; Score 5690; DB 11; Length 1188;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1071; Conservative 57; Mismatches 60; Indels 0; Gaps 0;

QY	1	MDLPRGLVAVALSMPGFTDTFNMTRKPRVPGSRTAFFGYTVQOHDISGNKWLIVGA	60
Db	1	MDLPRGLVAVALSMPGFTDTFNMTRKPRVPGSRTAFFGYTVQOHDISGNKWLIVGA	60
QY	61	ELETNCYQKTDGVYKCPVTHGNCCTKLNLGRVTLNVSEKRDNNRLGLSLATNPKNDSFLA	120
Db	61	PMETNGHQXTGVDYKCPVTHGNCCTKLNLGRVTLNVSEKRDNNRLGLSLATNPKNDSFLA	120

QY 121 CSPLWSHCEGSSYTTGMCSSRVNSNFRSKTVAPALQRCQTYMDIVIVLDGNSNIYPWVE 180
Db 121 CSPLWSHCEGSSYTTGMCSSRVNSNFRSKTVAPALQRCQTYMDIVIVLDGNSNIYPWVE 180
QY 181 VQHFLINILKKFYIGPGQIQGVVQYGEDVNHFHNDYRSVKDVEAASHIEQGGTET 240
Db 181 VQHFLINILKKFYIGPGQIQGVVQYGEDVNHFHNDYRSVKDVEAASHIEQGGTET 240
QY 241 RTAFGIEFARSAFQKGRKGAKKVMI VITDGESDSDPLEKVIQOESRDNVTRAVAVL 300
Db 241 RTAFGIEFARSAFQKGRKGAKKVMI VITDGESDSDPLEKVIQOESRDNVTRAVAVL 300
QY 301 GYNNRGINPEFLNEIKYIASDPDDKHFNFVNTDEAALKDVIDALGDRIFSLEGTNKNET 360
Db 301 GYNNRGINPEFLNEIKYIASDPDDKHFNFVNTDEAALKDVIDALGDRIFSLEGTNKNET 360
QY 361 SPGLEMSQTFSSHVVEDGVLGAVGYDMNGAVLKETSAGKVIPLRESYLKEPPEELKN 420
Db 361 SPGLEMSQTFSSHVVEDGVLGAVGYDMNGAVLKETSAGKVIPLRESYLKEPPEELKN 420
QY 421 HGAYLGTYTSSVSSRQGRVYVAGAPRNFHTGKVLFTMHNNRSLTIHQAMRGQQTGSYF 480
Db 421 HGAYLGTYTSSVSSRQGRVYVAGAPRNFHTGKVLFTMHNNRSLTIHQALRGEQIGSYF 480
QY 481 GSEITSDVIDGDGVTDLVLLGAPMYFNEGREGKVVYVELRQNRFYVNGTLKDSHSYQNA 540
Db 481 GSEITSDVNDNRVTDVLLGAPMYFSEGREGKVVYVNLQRNFYVNGTLKDSHSYQNA 540
QY 541 RFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKQITASELATG 600
Db 541 RFGSCIASVDLNQDSYNDVVGAPLEDNHRGAIYIFHGFQTNILKKPMQRIITASELAPG 600
QY 601 LQFPGSIHQDLNEDGLIDLAVGALGNVILWSPVQINASLHFEPSKINIFHRDCK 660
Db 601 LQFPGSIHQDLNEDGLIDLAVGALGNVILWSPVQINASLHFEPSKINIFHRDCK 660
QY 661 RSGRDATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYTPRAHLDGEGDRFNRVAV 720
Db 661 RSGRDATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYTPRAHLDGEGQFTNRVAV 720
QY 721 LLSGQELCBRINFHVLDTADYKPVTFSEYSELEDDHGMPLDDGWPPTLRYSVFPWNG 780
Db 721 LLSGQELCBRINFHVLDTADYKPVTFSEYSELEDDHGMPLDNGWPTLRYSVFPWNG 780
QY 781 CNEDEHCVPLVDARSDDLPTAMEYCORVLKPKAQDCSAYTSLSDTTVFIESTRQVAV 840
Db 781 CNEDEHCVPLVDARSDDLPTAMEYCORVLKPKAQDCSSYTLSDTTVFIESTRQVAV 840
QY 841 EATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNEERLQKQVCNVSYPFF 900
Db 841 EATLENRGENAYSAVLNISQSANLQFASLIQKEDSDNSIECVNEERLHKVKVCNVSYPFF 900
QY 901 RAKAKVAFRLDSEFSKSI FLHLEIETLAAGSDSNRSTKEDNVAPLRFHLKYEADVLFT 960
Db 901 RAKAKVAFRLDSEFSKSI FLHLEIETLAAGSDSHEQDSTADNTALLRFLHLYEADVLFT 960
QY 961 RSSLSHVEKINSLSERYDGGIPFPFCIFRQNLGLPFIHGMKMTIPIATRSNRL 1020
Db 961 RSSLSHVEKANSLESYDGGIPFNCVQNLGFPFIHGMKMTIPIATRSNRL 1020
QY 1021 KLRDPLTDANTSCNINWNSTEYRTPVEEDLRAPQNLHNSDVVSNICNIRLVNQE 1080
Db 1021 MLRDFDTQGNSTSCNINWNSTEYRTPTEEDLSHAPQNLHNSDVVSIICNIRLVNQE 1080
QY 1081 NPHLGNLWLSLKALKYKSMKIMVNAALQRFHSHFFIFREDDPSRQFEFISKOEDQV 1140
Db 1081 SFYLVGNLWLSLKALKYKSMKIMVNAALQRFHSHFFIFREDDPSRQFEFISKOEDQV 1140
QY 1141 PIWIIIVGSLGLGLLALLVLAALRKLGFPSARRRPEGLDPTPKVLE 1188
Db 1141 PIWIIIVGSLGLGLLALLVLAALRKLGFPSARRRPEGLDPTPKVLE 1188

RESULT 2
Q8WY18 PRELIMINARY; PRT; 823 AA.
AC Q8WY18;
DT 01-MAR-2002 (TREMREL. 20, Created)
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE MSTP018.
GN MST018.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
Gao R.L., Qiang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF111799; AAL39001.1; -.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro: IPR000413; Integrin_alpha.
DR Pfam: PF01839; FG-GAP; 3.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 4.
SQ SEQUENCE 823 AA; 92672 MW; DEAE78079DCD4925 CRC64;
Query Match 68.6%; Score 4272; DB 4; Length 823;
Best Local Similarity 99.6%; Pred. No. 2.9e-288;
Matches 820; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 366 MSQTGFSSHVVEDGVLGAVGYDMNGAVLKETSAGKVIPLRESYLKEPPEELKNHAYL 425
Db 1 MSQTGFSSHVVEDGVLGAVGYDMNGAVLKETSAGKVIPLRESYLKEPPEELKNHAYL 60
QY 426 GYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNNRSLTIHQAMRGQQTGSYFSEIT 485
Db 61 GYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNNRSLTIHQAMRGQQTGSYFSEIT 120
QY 486 SVDIDGDGVTDLVLLGAPMYFNEGREGKVVYVELRQNRFYVNGTLKDSHSYQNAFESS 545
Db 121 SVDIDGDGVTDLVLLGAPMYFNEGREGKVVYVELRQNRFYVNGTLKDSHSYQNAFESS 180
QY 546 IASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKQITASELATGLOYFG 605
Db 181 IASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKQITASELATGLOYFG 240
QY 606 CSIHQGLDNLNEDGLIDLAVGALGNVILWSPVQINASLHFEPSKINIFHRDCKRSGRD 665
Db 241 CSIHQGLDNLNEDGLIDLAVGALGNVILWSPVQINASLHFEPSKINIFHRDCKRSGRD 300
QY 666 ATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYTPRAHLDGEGDRFNRVAVLSSG 725
Db 301 ATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYTPRAHLDGEGDRFNRVAVLSSG 360
QY 726 QELCERINFHVLDTADYKPVTFSEYSELEDDHGMPLDDGWPPTLRYSVFPWNGCNEDE 785
Db 361 QELCERINFHVLDTADYKPVTFSEYSELEDDHGMPLDDGWPPTLRYSVFPWNGCNEDE 420
QY 786 HCVPLVDARSDDLPTAMEYCORVLKPKAQDCSAYTSLSDTTVFIESTRQVAVEATLE 845
Db 421 HCVPLVDARSDDLPTAMEYCORVLKPKAQDCSAYTSLSDTTVFIESTRQVAVEATLE 480
QY 846 NRGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNEERLQKQVCNVSYPFFRAKAK 905
Db 481 NRGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNEERLQKQVCNVSYPFFRAKAK 540
QY 906 VAFRLDSEFSKSI FLHLEIETLAAGSDSNRSTKEDNVAPLRFHLKYEADVLFTRSSSL 965

Db 541 VAFRLDFEFSKIFLHLLELAAGSDSNERDSTKEDNVAFLRHLKYEADVLFRSSSL 600
QY 966 SHYEVKLNLSRLRYDGLGPPFCIFRIQNLGLPPIHGMKKITIPATRSNRLKLKRD 1025
Db 601 SHYEVKLNLSRLRYDGLGPPFCIFRIQNLGLPPIHGMKKITIPATRSNRLKLKRD 660
QY 1026 LTDEANTSCNIGWNSTEYRPTVEEDLRRAPQLNHSNDVSNINRLVNPQEIHFLL 1085
Db 661 LTDEANTSCNIGWNSTEYRPTVEEDLRRAPQLNHSNDVSNINRLVNPQEIHFLL 720
QY 1086 GNLWLSLKALKYKSKIMWNAALQRFHSPPTFREEDPSRQIEFBSKQEDQVPIWII 1145
Db 721 GNLWLSLKALKYKSKIMWNAALQRFHSPPTFREEDPSRQIEFBSKQEDQVPIWII 780
QY 1146 VGSTLGGLLLLALLVLAALRKLGFRRARRRREPGLDPTPKVLE 1188
Db 781 VGSTLGGLLLLALLVLAALRKLGFRRARRRREPGLDPTPKVLE 823

RESULT 3
Q8CE84
ID Q8CE84 PRELIMINARY; PRT; 823 AA.
AC Q8CE84;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE MSTP018 homolog.
GN 4732459H24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK028821; BAC26137.1; -.
DR MGD; MGI:2442114; 4732459H24RIK.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; P:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 3.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int alpha; 4.
SQ SEQUENCE 823 AA; 92364 MW; A330236324A0E089 CRC64;

Query Match 61.7%; Score 3840; DB 11; Length 823;
Best Local Similarity 87.8%; Pred. No. 3.4e-258;
Matches 723; Conservative 51; Mismatches 49; Indels 0; Gaps 0;

QY 366 MSQTGFSSHVVEDGVLGAVGAYDNGAVLKETSAGKVIPLRESYLKFEPEELKNHGAYL 425
Db 1 MSQTGFSSHVVEDGILLGAVGAYDNGAVLKETSAGKVIPLRESYLKFEPEELKNHAYL 60
QY 426 GYTVTSVSSRQGRVYVAGAPRFNHTGKVLFTMHNKSLTIHCAMRGQOIGSYFGSEIT 485
Db 61 GYTVTSVSSRQGRVYVAGAPRFNHTGKVLFSMHNKSLTIHCAALGEQIGSYFGSEIT 120
QY 486 SVDIDGQVTDVLLVGAPMFMNEGREGKVVYVYELQRNFYNGTLKDSHSYQNAFGSS 545
Db 121 SVDVNDVRDVTLLVGAPMFMFSEGRGKVVYVYELQRNFYNGTLKDSHSYQNAFGSC 180
QY 546 IASVRDLNQSNDVYVVGAPLEDNHAGAIYIFHGFRGSIKTPKQRTASLQYFG 605
Db 181 IASVDLQNSNDVYVVGAPLEDNHRGAIYIFHGFRGSIKTPKQRTASLQYFG 240

QY 606 CSHGOLDLNEDGLIDLAVGALGNVILMSRPVVQINASLHFPSKINIFHRCKRSRD 665
Db 241 CSHGOLDLNEDGLVDLAVGALGNVILMSRPVVQINASLHFPSKINIFHRCKRSRD 300
QY 666 ATCLAAFLCPTPIFLAPHFQTTVGRYNATMDERRYPTRAHLDEGDRTRNAVLLSSG 725
Db 301 ATCLAAFLCPTPIFLAPHFQTTVGRYNATMDERRYPTRAHLDEGDRTRNAVLLSSG 360
QY 726 QELCERINFLVLDADYVVKPVPFVSVEYSLEDPDHGPMLDDGWPTTLTVSVVPFNGCNEDE 785
Db 361 QELCERINFLVLDADYVVKPVPFVSVEYSLEDPDHGPMLDDGWPTTLTVSVVPFNGCNEDE 420
QY 786 HCVPLDVLDAESDLPTAMEYCORVURKPAQDCSAYTSLSEDTTTFITESTRQVAVETLE 845
Db 421 HCVPLDVLDAESDLPTAMEYCORVURKPAQDCSAYTSLSEDTTTFITESTRQVAVETLE 480
QY 846 NEGENAYSTVLNISOSANLOFASLIQKEDSDGSIETCVNERELQKQVNCVSPFFRAKAK 905
Db 481 NEGENAYSTVLNISOSANLOFASLIQKEDSDGSIETCVNERELQKQVNCVSPFFRAKAK 540
QY 906 VAFRLDFEFSKIFLHLLELAAGSDSNERDSTKEDNVAFLRHLKYEADVLFRSSSL 965
Db 541 VAFRLDFEFSKIFLHLLELAAGSDSNERDSTKEDNVAFLRHLKYEADVLFRSSSL 600
QY 966 SHYEVKLNLSRLRYDGLGPPFCIFRIQNLGLPPIHGMKKITIPATRSNRLKLKRD 1025
Db 601 SHYEVKLNLSRLRYDGLGPPFCIFRIQNLGLPPIHGMKKITIPATRSNRLKLKRD 660
QY 1026 LTDEANTSCNIGWNSTEYRPTVEEDLRRAPQLNHSNDVSNINRLVNPQEIHFLL 1085
Db 661 LTDEANTSCNIGWNSTEYRPTVEEDLRRAPQLNHSNDVSNINRLVNPQEIHFLL 720
QY 1086 GNLWLSLKALKYKSKIMWNAALQRFHSPPTFREEDPSRQIEFBSKQEDQVPIWII 1145
Db 721 GNLWLSLKALKYKSKIMWNAALQRFHSPPTFREEDPSRQIEFBSKQEDQVPIWII 780
QY 1146 VGSTLGGLLLLALLVLAALRKLGFRRARRRREPGLDPTPKVLE 1188
Db 781 VGSTLGGLLLLALLVLAALRKLGFRRARRRREPGLDPTPKVLE 823

RESULT 4
O42094
ID O42094 PRELIMINARY; PRT; 1171 AA.
AC O42094;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE ALPHA integrin.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gizzard;
RX MEDLINE=97476270; PubMed=9334246;
RA Obata H., Hayaishi K., Nishida W., Momiyama T., Uchida A., Ochi T.,
RA Sobue K.;
RA "Smooth muscle cell phenotype-dependent transcriptional regulation of
RT the alpha1 integrin gene."
RL J. Biol. Chem. 272:26643-26651 (1997).
DR EMBL; AB000470; BAA23160.1; -.
DR EMBL; AB000471; BAA23161.1; -.
DR PIR; A56348; A55348.
DR HSP; P17301; IAOX.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.

Db 8 LLOGWVLASCHG-----SNLDVEEP-IVFREDAASFGQTVQF-----GGSLRVVGLAPLEAVA 59
QY 67 YQKTDGVYKCPVIHGNCTKLNGLRVTLNSVSEKDNMRLGLSLATNPKDNSFLACSPWS 126
Db 60 VNQGTGLYDCAPATGMCQPIVL-RSPLEAV-----NMSLGLSLVTA TNNAQLLACGPTAQ 113
QY 127 HECGSSYTTGMCSSRVNSNFRFSKTVAPALQRC-QTYMDIVIVLDGNSNI--YPMWVEVQH 183
Db 114 RACVKNMYAKGSCLLGSSQLQIQA VPA SMCPECRQEMDIAFLIDGSSINQORDFAQMKD 173
QY 184 FLINILKFFYIGPQIQVGVVOVGEDVVEHFLNDYRSVKDVVEAASHIIFQRGTTETRTA 243
Db 174 FVKALMGEF--ASTSTLPSMGSVNLKTHFTTFFTKNILDPOQLVDPIVQLQGL-TYTA 230
QY 244 FGIEFARSEAF--QKGRGKAKVMIVITDGEHSDSP-DLEKVIQOQSERDNVRYAVAVL 300
Db 231 TGI RTWEEELFHSGKNGSRKSAKILLVITDQKVRDPLEYSDVIPAADKAGIIRYALGVG 290
QY 301 GYNNRGINPETFLNEIKYIASDDDKHFFNVTDKALDIDVADGDRIFSLGEGT-NKNE 359
Db 291 DAFQE-----PTALKELNTIGSAPPQDHVFKVGNFAALRSIQRLQEKIFAIEGTQSRSS 345
QY 360 TSFGLMSQTRGSHVVDGVLGAVCAVDWNGAVLKETSAGKVP--LRESYLKFFPBE 417
Db 346 SSFQHEMSQGRSSALTS DGPLGAVGSPFSWSGAF-----LYPNTREPTFINMSQEN 398
QY 418 LKNHGYLGYTVTSVSSKQGRVYVAGAPRFNHTGKVLFTMNNRSLTIHQAMRQQOIG 477
Db 399 VDMRDSYLGYS-TAVAFWKGVHSLILGAPRHQHTGKVIFT-QEARNRPKSEVRQTQIG 456
QY 478 SYFGSEITSDVDIGDGTVDVLLGAPMNEGRERKGVVYELR--QNRFPVYNGTLKDSH 535
Db 457 SYFGASLCSVDVRDSTDLVIGAPHYEQTR-GQGVSVFPVPGVGRWQCEATLHGBO 515
QY 536 SYQNARFSGSIASVRDLNDSYNDVVVGAPLENNHAGIYIPIRG-PRGSILKTPQRIITA 594
Db 516 GHPWGRFGVALVIGDVNGDNLADVAIGAPGEESESGAVIYIPIRGASRLIMPSPSRVTG 575
QY 595 SELATGLQVFGSGIHGOLDNEDGLIDLAGVAGNALWSRPVGVGINASLHFEFSKINI 654
Db 576 SOLSLRLQVFGSLSGQDLTQDGLVLAVGAGHVLRLSLPLKVELSIRFAPMEVAK 635
QY 655 PHRCKRSRDRATCL---AAFLCTPIFIAPHQTTTGV-IRYNATMDERRYTPRAHLDE 710
Db 636 AVYQCWE--RTPTVLEAGEATVCLTVHKGSPDLLGNVGSVRYDLALDPLGRILSRAIFDE 693
QY 711 GGDREFTNRAVLSSGOELCERLNFHVL-DADYVVKPVTFSVEYSL-----EDPDHGMID 764
Db 694 TKNCCTLTKGKTLGLGDH-CEIVKLLLPDCEVDVSPILRLNFSLVRSASPRNLHPVLA 752
QY 765 DGWPTTLRVSVFPMWNCNEDEHCVPLDVLDA RSDLP TAMEYQORVLRKPAQDCSAYTLSP 824
Db 753 VGSQDHITASLPFEKNCKQELLCEGL-----GISENF 785
QY 825 DTTVFIIESTQRVAVEATLENRGNAYSTVNISQSANLQFASLI---QKEDSDGSIEC 881
Db 786 SGLQVLVVGSGSPELTVTVVMNEGDSYGLVKFYYPAGLSYRRVTGTQCPHQYPLRLAC 845
QY 882 VNE---ERRLOKQCNVSPYFFRAKAVARLDSFSSIFL-HHLELELAAGSDSNRDR 937
Db 846 EAEPAQEDLRSSSCSINHPIFREGAKTFTMITFDVSYKAFGLDRLLLRKASSENKPD 905
QY 938 STKENDVAPLPHLYEADVLFTRSSLSHSEVKLNSSLERYDGI GPPFCIFRIQNLGL 997
Db 906 TNK--TAFQELSPVKYTVVTLISQEDSTNH-VNFSSS---HGRRQPAAHRYVNNLSP 959
QY 998 FPIHGMKMTIPIATRSNGNLLKLRLDFTDEANTSCNIMGNSTETPTPVEE-----DL 1052
Db 960 LKL-AVRVNFVWVPL-----LNGVAVWVTLSSPAQGVSCVQMKPQNFDFLTQI 1009
QY 1053 RRAPOLHNSDVVSINCNIRLVNQ-EINPHILGNL---WLRL---KALKYKSMKIMV 1105
Db 1010 QRRSVLDCSIADCLHFRCDIPSLDIQDELDFILRGNLSFGWVSQTLQEKVLLVSEAITF 1069

QY 1106 NAALQRFHSPFFIFREEDPSRQIEFEISKQEDWQV--PIWIVGSTLGGILLALLVIAL 1163
Db 1070 DTSVYSLPQGEAFRLR-----AQVETTL---EYVVVEPIFIVAGSSVGGILLALLITVVL 1122
QY 1164 RKLGFER 1170
Db 1123 YKLGFEK 1129
RESULT 6
Q9QXH4
ID Q9QXH4 PRELIMINARY; PRT; 1169 AA.
AC Q9QXH4;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Leukocyte adhesion glycoprotein p150,95 alpha integrin subunit.
GN ITGAX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang X., Gorski K., Tong C., Rattis F.-M., Tseng S.-Y., Pardoll D.,
RA Tsuchiya H.;
RT "Isolation of Genes Selectively Expressed by Dendritic Cells.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF211864; AAF23492.1; --
DR HSSP; P11215; 1BHQ.
DR MGD; MGI:96609; Itgax.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
KW Integrin.
SQ SEQUENCE 1169 AA; 129150 MW; C616412033C219A6 CRC64;
Query Match 18.6%; Score 1160; DB 11; Length 1169;
Best Local Similarity 29.5%; Pred. No. 2.3e-71;
Matches 372; Conservative 215; Mismatches 498; Indels 174; Gaps 49;
QY 9 VAWALSLPWPGTDT--FMOTRKPRVPGSTAF-----FGYTVQQHDISGNKWLVVGA 60
Db 6 IAFLLLL--GVSCLGFLNDAEK-----LTFHMDGAEFGHGVLYDSS---WVVVGA 53
QY 61 PLETNGYQKTVGVKCPVIHGNCTKLNGLRVTLNSVSEKDNMRLGLSLATNPKDNSFLA 120
Db 54 PKEIKATNQIGGLYKCYHTGNCETPSL-----QVPPEAVNTISLGLSLAATNPSWLLA 107
QY 121 CSPLWSHECGSSYYTTCMSRVNSNFRFSKTVAPALQRC-QTYMDIVIVLDGNSNI--YP 177
Db 108 CGPTVHTTCRENIYLTGLCFLLSSSFQSQNFPTAQCECPQODIVFLIDGSGSISSTD 167
QY 178 WVEVQHFLINILKFFYIGPQIQVGVVOVGEDVVEHFLNDYRSVKDVVEAASHIEQGG 237
Db 168 FEXMLDFVKVAMSQ--LQRESTRFSLMKQFSDYFRVHFTFNNFISTSSPLSLGSRVLRG 225
QY 238 TETRTAFGIFFARSEAF--QKGRGKAKVMIVITDGEH-DSFDLEKVTQOQSERDNVTR 294
Db 226 -YTTASAIKHVITELTTSQSGARQATKVLIVITDGRKQGNLSYDVSIVPMAEASIIIR 284

Qy	874	DSGSIETCNEER-BLQKQVCNVSYPFRKAKAVP-----RLDSFESKIFUHLLEI	925
Db	834	RTRSS--CGDRSGATNRTCSIDLPPVRSQTTFGLGTFRVKWKWDNDSS-----NRMEI	886
Qy	926	ELAASDSN-EDSTKEDNVAFLPHLKYADVLTFRSSLSHYEVKLNSSLERYDGTGP	984
Db	887	MITANSNDNGNSMDEVRSVPQFAV--DLAISLVAEDSVTY---NMFSLE---DRGP	937
Qy	985	-PFCIFRIONGLPPIHGMMKTIPIATRSNGNRLKLRLDFLTDEANTS-----	1033
Db	938	KPLNTYKVENGR---KDLPSVILTLPQTPHVLITHTESMHEVHSHFISSYHQIIM	994
Qy	1034	CNIWGSTERYPTVEEDLRRAPQLNHSNSDVVSINCN-IRLVNQNINPHLLGNLWLS	1092
Db	995	CLLNKHLFFSP-----ELSAVOQVRSTGWSLREVECSQFDLNMKSAVHNLTDARLQN	1049
Qy	1093	LKALKYK-----SMKIMVNAALQROPHSPFIFREEDPSR-QIEFELSKQED	1137
Db	1050	VKEYESKSYFPRKDNVFSISAEILNYSNLNQTSELKY---NPHRSQTEVKV---E	1102
Qy	1138	WQVP---IWIIVGSTLGLLLALLVLALRLKLGFFRSAR	1173
Db	1103	FVVPSSLMLIVCTGAVGGFFLIILFLLLKCGFFXRNR	1141
RESULT 9			
Q9J130			
ID	Q9J130	PRELIMINARY; PRT; 1151 AA.	
AC	Q9J130	Q9J130; 2000 (TEMBLrel. 15, Created)	
DT	01-OCT-2000	(TEMBLrel. 15, Last sequence update)	
DT	01-OCT-2000	(TEMBLrel. 15, Last sequence update)	
DT	01-JUN-2003	(TEMBLrel. 24, Last annotation update)	
DE	Integrin beta 2 alpha subunit.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=101116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Fathallah D.M. Sr., Zerrila K. Jr.;		
RT	"Cloning of the rat CD11b cDNA sequence."		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF268593; AAF81280.1; -		
DR	HGSP; P11215; IAHQ.		
DR	GO; GO:0008305; C:integrin complex; IEA.		
DR	GO; GO:0004895; F:cell adhesion receptor activity; IEA.		
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.		
DR	InterPro; IPR000413; integrin_alpha.		
DR	InterPro; IPR002035; VNF_A.		
DR	Pfam; PF01839; FG-GAP; 3.		
DR	Pfam; PF00357; integrin_A; 1.		
DR	Pfam; PF00092; vwa; 1.		
DR	PRINTS; PR01185; INTEGRINA.		
DR	PRINTS; PR00453; VWFADOMAIN.		
DR	SMART; SM00191; Intc_alpha; 5.		
DR	SMART; SM00327; VWA; 1.		
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.		
DR	PROSITE; PS0234; VWF_A; 1.		
SQ	SEQUENCE	1151 AA; 126943 MW; 8F785695D4074CA5 CRC64;	
Query Match			
Best Local Similarity			
Matches 357; Conservative 203; Mismatches 494; Indels 185; Gaps 49;			
Qy	7	LVAWALSMPGFTDTFNMDCRXPVPGSRTPAFGYTVQOHDISGNKMLVVGAPLETNG	66
Db	5	VLLATVLTCHG-----FNLDTENPMTFOENASG-FGQSVIQ---LGETRVVVAAPQEVKA	56
Qy	67	YQRTGADVYKCPVHGNCTKLNGLRVTLSNVSERKDNMELGLSLATNPKNSEFLACPLWS	126
Db	57	VNQTGALYQCDYSTNCRCDPIPL-----QVPPEAVNMSLGLSLAATVPPQLLAGPTVH	110

OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TISSUE=Spleen;	
RA	Ma R.Z., Teuscher C.;	
RL	Submitted (May-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF065901; AAD25884.1; -	
DR	HSSP; P20701; 1LEA.	
DR	MGI; 96606; Itgal.	
DR	GO; GO:0008305; C:integrin complex; IEA.	
DR	GO; GO:0004895; F:cell adhesion receptor activity; IEA.	
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.	
DR	InterPro; IPR000413; Integrin_alpha.	
DR	InterPro; IPR002035; VWF_A.	
DR	Pfam; PF01839; FG-GAP; 3.	
DR	Pfam; PF00357; integrin_A; 1.	
DR	Pfam; PF00092; vwa; 1.	
DR	PRINTS; PR01185; INTEGRINA.	
DR	PRINTS; PR00453; VWFADOMAIN.	
DR	SMART; SM00191; Int_alpha; 5.	
DR	SMART; SM00327; VWA; 1.	
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.	
DR	PROSITE; PS0234; VWEA; 1.	
SQ	SEQUENCE 1160 AA; 128127 MW; A33C531B139FlFAD CRC64;	
Query Match		
Best Local Similarity 28.1%; Pred. No. 1.3e-64;		
Matches 358; Conservative 204; Mismatches 451; Indels 262; Gaps 55;		
QY	4	PRGLVAVAWL--SLWPGFTTFNMDTRKPRVIFGSRITAFPGYTVQQHDISGNKWLVGAP 61
DB	8	PRLLGLQLFAKAW-----SYNLDRTPQSFQAQGRHEGYQLQIE-----DGVVVGAP 58
QY	62	LETNGYQKTDGVYKCPVIHGNCKKLNLRVTLNVSRKDNMRIGLSLATNPKDNSFLAC 121
DB	59	GEGD---NTGLVHCRITSFECQPVSLH-----GSNHTSKYLGMTLATDAAGKSLAC 108
QY	122	SPLWSHECGSSYYTGMCSRNVGNFRSKTV-----APALQRC-QTYMDIIVILDGSN 173
DB	109	DGGLSRTCDQNTYLSGLC-----YLPQSLGFMQLQNRPAYQECMKGVLDVLFDFGSQ 162
QY	174	SI--YPAVEVQHFLINILKKFYIGPGQIQGVQVQGEDVHVEHLNDRSVK-----DVV 226
DB	163	SLDRKDFEKLIFEMKQWRK--LSNTSYQFAAQVQFSTDCRTEFTFLDY--VKQKNPDVL 218
QY	227	EAASHIEQRGGE-----TRTAFGIEFARSEAF--QKGRGAKKVMIVITDGESHSDPD 279
DB	219	-----LGSVQPMFLTNTFRAINVVAHVFKESGARPDATKVLVITDGEASDKGN 270
QY	280	LEKVIQOSERNVTRYAVAVLGYNNRGINPETFNLNFIKVIASDPDDKHFNVTDE-AAL 338
DB	271	I-----SAAHDITRYIIGIKHP--VSVQKQKTLH---IFASEPVVEE-FVKILDTEFL 318
QY	339	KDIYDALGDRIFSLEGNKNE-TSFGLEMSQTGSPSSHVEDGVLLGAVGAYDNMGAVLKE 397
DB	319	KDLFTDQRRYIALEGNRQDLTSFNMLSSGSLASDLKSHAVVGAVGAKW----- 371
QY	398	TSAGKVIPLRE-----SYLKEFPPELKNHGLAYLTYTTSVVSSRQGRVYVAGAPRNFHTG 452
DB	372	--AGGFLDLREDLQAGATFVGQEPLTSDVRGGLGYTVVAMTSSRSPRLAAGAPRYQHV 429
QY	453	KVILFTMHN-----NRSLTIHQMRGQIGSYFSEITSVDIDGGVTDVLLVGAPMYFN 507
DB	430	QVLLFOAPEAGGRWNOT---QKLEGTQIGSYFGELCSVDLDQGEAELLILGAPLFG 485
QY	508	EGERGKVVYVELQRNRYVNGTLKDSHYQNARFGSSIASVRDLNQDSYNDVNVVVGAPLE 567
DB	486	EQR-CGRVTTYQRROSLEFEMVSELQCDPGYPLGRFGAITALTDINGRLTDVAVGAPLE 544
QY	568	DNHAGAIYIFHGFRGSILKTPKQRTIASELATGLQYFCGSHQGLDLNEDGLIDLAVGAL 627
DB	545	EQ--GAVIFNGKPGGLSPQSRIQGAQVFGRFWGRSHGVKDLGGDRLANVVGPE 602
RESULT 12		
Q98TF0	Q98TF0	PRELIMINARY; PRT; 1187 AA.
AC	Q98TF0;	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	CD11-2.	
GN	CIA2.	
OS	Cyprinus carpio (Common carp).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;	
OC	Cyprinidae; Cyprinus.	
OX	NCBI_TaxID=7962;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Peritoneal exudate cells;	
RA	Kimura M., Fujiki K., Nakao M.;	
RT	"Molecular cloning of a leukocyte integrin from the common carp.;"	
RL	submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AB048537; BAB39135.1; -	
DR	HSSP; P20701; 1LEA.	
DR	GO; GO:0008305; C:integrin complex; IEA.	
DR	GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.	
DR	GO; GO:0004895; F:cell adhesion receptor activity; IEA.	
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.	
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.	
DR	InterPro; IPR001969; Asprotease_AS.	

[illegible]

Tue Sep 21 14:48:57 2004

[illegible]

Search completed: September 21, 2004, 13:03:42
Job time : 187.196 secs

Title: US-09-980-403-2
 Perfect score: 6224
 Sequence: 1 MDLPRGLVAVALSMPGFT.....FRSARRREPGLDPTPKVLE 1188
 Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Searched: 1586107 seqs, 282547505 residues
 Total number of hits satisfying chosen parameters: 1586107
 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : A_Geneseq_29Jan04:.*
 1: Geneseqp1980s:.*
 2: Geneseqp1990s:.*
 3: Geneseqp2000s:.*
 4: Geneseqp2001s:.*
 5: Geneseqp2002s:.*
 6: Geneseqp2003as:.*
 7: Geneseqp2003bs:.*
 8: Geneseqp2004s:.*

Run on: September 21, 2004, 12:43:42 ; Search time 226.483 Seconds
 (without alignments)
 1482.081 Million cell updates/sec
 OM protein - protein search, using sw model
 Copyright (c) 1993 - 2004 Compugen Ltd.
 GenCore version 5.1.1.6

Run on: September 21, 2004, 12:43:42 ; Search time 226.483 Seconds
 (without alignments)
 1482.081 Million cell updates/sec
 OM protein - protein search, using sw model
 Copyright (c) 1993 - 2004 Compugen Ltd.
 GenCore version 5.1.1.6

Run on: September 21, 2004, 12:43:42 ; Search time 226.483 Seconds
 (without alignments)
 1482.081 Million cell updates/sec
 OM protein - protein search, using sw model
 Copyright (c) 1993 - 2004 Compugen Ltd.
 GenCore version 5.1.1.6

Run on: September 21, 2004, 12:43:42 ; Search time 226.483 Seconds
 (without alignments)
 1482.081 Million cell updates/sec
 OM protein - protein search, using sw model
 Copyright (c) 1993 - 2004 Compugen Ltd.
 GenCore version 5.1.1.6

Run on: September 21, 2004, 12:43:42 ; Search time 226.483 Seconds
 (without alignments)
 1482.081 Million cell updates/sec
 OM protein - protein search, using sw model
 Copyright (c) 1993 - 2004 Compugen Ltd.
 GenCore version 5.1.1.6

26 2401.5 38.6 1152 4 AAB64657 Human sec
 27 2375.5 38.2 1167 4 AAB64584 Human sec
 28 2375.5 38.2 1167 6 ABP99490 Human sec
 29 2375.5 38.2 1167 6 ABR00964 Human gen
 30 2375.5 38.2 1167 6 ADA44026 Human sec
 31 2373.5 38.1 1152 4 AAB64658 Human sec
 32 2319 37.3 1132 3 AAB64658 Human int
 33 2166.5 34.8 1179 5 ABB90759 Human int
 34 2166.5 34.8 1179 5 ABB90759 Human int
 35 2166.5 34.8 1179 5 ABB90759 Human int
 36 2159 34.7 1151 8 ADE86584 Human int
 37 2149.5 34.5 1180 5 ABB90788 Human int
 38 2149.5 34.5 1180 5 ABB90788 Human int
 39 2149.5 34.5 1180 6 ABU54495 Mouse tum
 40 2093 33.6 1183 2 AAY07728 Armenian
 41 1863 29.9 1181 6 ABU03548 Angiogene
 42 1863 29.9 1181 6 ABR59703 Human VIA
 43 1863 29.9 1181 6 AAG79775 Alpha2 in
 44 1863 29.9 1181 6 ABU03616 Human exp
 45 1863 29.9 1181 6 ABU03614 Human exp

ALIGNMENTS

RESULT 1
 AAB30929
 ID AAB30929 standard; protein; 1188 AA.
 XX AC AAB30929;
 XX DT 02-APR-2001 (first entry)
 XX DE Amino acid sequence of a human alpha1 integrin chain.
 XX KW Human; integrin; alpha1 subunit; fibroblast; muscle cell; chondrocyte;
 KW osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;
 KW wound healing; trauma; rheumatoid arthritis; osteoarthritis;
 KW osteoporosis; cartilage damage; bone damage; cartilage.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /note= "signal peptide"
 FT Region 951..972
 FT /note= "leucine zipper"
 FT Domain 1142..1164
 FT /note= "transmembrane domain"
 FT WO200075187-A1.
 FT 14-DEC-2000.
 FT 31-MAY-2000; 2000WO-SE001135.
 FT 03-JUN-1999; 99SE-00002056.
 FT (ACTI-) ACTIVE BIOTECH AB.
 FT Gullberg D;
 FT WPI; 2001-071061/08.
 FT N-PSDB; AAC86871.
 XX Integrin subunit alpha 11 or integrin heterodimer comprising subunit
 PT alpha 11 in association with subunit beta, useful for treating muscle
 PT dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis.
 XX Disclosure; Fig 2a-c; 79pp; English.
 XX The present sequence represents a human integrin subunit, designated
 CC alpha11. The alpha11 polynucleotide and polypeptide are useful as markers

CC of cell target molecules, such as fibroblasts, muscle cells,
 CC chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.
 CC They are also used for determining the differential-stage of cells during
 CC differentiation, development in pathological conditions, in tissue
 CC regeneration, in transplantation or in therapeutic and physiological
 CC repair of tissues. The pathological conditions involving subunit alpha1
 CC are selected from damage of cells, muscle dystrophy, fibrosis, wound
 CC healing, trauma, rheumatoid arthritis, osteoarthritis and osteoporosis,
 CC damage of cartilage and bone, and cartilage and bone diseases. The
 CC polypeptide is useful for detecting the formation of cartilage during
 CC embryonic development, for detecting physiological therapeutic repair of
 CC cartilage and muscle, for selection and analysis, or for sorting,
 CC isolating or purification of chondrocytes and muscle cells, for detecting
 CC regeneration of cartilage or chondrocytes during transplantation of
 CC cartilage or chondrocytes during transplantation of cartilage or
 CC chondrocytes, respectively, or of muscle or muscle cells during
 CC transplantation of muscle or muscle cells, respectively, and for studies
 CC of differentiation or chondrocytes or muscle cells
 XX
 XX Sequence 1188 AA;

Query Match 100.0%; Score 6224; DB 4; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLPRGLVAVALS LWPGETDTFNMDTRKPRVPGSRTPAFGTYVQOHDISGNKWL VVGA 60
 DB 1 MDLPRGLVAVALS LWPGETDTFNMDTRKPRVPGSRTPAFGTYVQOHDISGNKWL VVGA 60

QY 61 PLETTYGYOKTGVYKCPVTHGCTKLNIGRVTL SNVSRKONMRLGLSLATNPKDNSFLA 120
 DB 61 PLETTYGYOKTGVYKCPVTHGCTKLNIGRVTL SNVSRKONMRLGLSLATNPKDNSFLA 120

QY 121 CSPLMSHSCGSSYYTTCMCSRNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSNYPWVE 180
 DB 121 CSPLMSHSCGSSYYTTCMCSRNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSNYPWVE 180

QY 181 VQHFLINILKFKYIGPQIQGVQVQGVGDVHFEHFLNDRSVKDVVEAAASHTEQRGGTET 240
 DB 181 VQHFLINILKFKYIGPQIQGVQVQGVGDVHFEHFLNDRSVKDVVEAAASHTEQRGGTET 240

QY 241 RTAFGLEPAREAFQGGKRGKAKVMIVITDGHSDSPLEKVIQOSERDNTRYAVAVL 300
 DB 241 RTAFGLEPAREAFQGGKRGKAKVMIVITDGHSDSPLEKVIQOSERDNTRYAVAVL 300

QY 301 GYNRRGINPETFLNFKIYASDPDKHFNVTDEAALKDIDVALGDRIFSLEGTNKNET 360
 DB 301 GYNRRGINPETFLNFKIYASDPDKHFNVTDEAALKDIDVALGDRIFSLEGTNKNET 360

QY 361 SFGLEMSQTFSSSHVVEDGVLLGAVGAYDMNGAVLKETSGAKVPIPLRESYLKEFPPELKN 420
 DB 361 SFGLEMSQTFSSSHVVEDGVLLGAVGAYDMNGAVLKETSGAKVPIPLRESYLKEFPPELKN 420

QY 421 HGAYLGYTVTSVSSRQGRVYVAGAPREHNTGKVILFTMHNRSLSITHQAMRQQQGSYF 480
 DB 421 HGAYLGYTVTSVSSRQGRVYVAGAPREHNTGKVILFTMHNRSLSITHQAMRQQQGSYF 480

QY 481 GSEITSDVIDGDGTVDLLVGAPMYFNEGRGKGVVYVELQRNFRVYNGTLKDSHYSYQNA 540
 DB 481 GSEITSDVIDGDGTVDLLVGAPMYFNEGRGKGVVYVELQRNFRVYNGTLKDSHYSYQNA 540

QY 541 RFGSSIASVRDLNODSNVDVVGAPLEDNAGAIYIPHGFRGSLKTPKQRIITASELATG 600
 DB 541 RFGSSIASVRDLNODSNVDVVGAPLEDNAGAIYIPHGFRGSLKTPKQRIITASELATG 600

QY 601 LOYFGCSIHGOLDLNEGLDIDLAVGALGNVILWSRPVQINASLHPEPSKINIFHRDCK 660
 DB 601 LOYFGCSIHGOLDLNEGLDIDLAVGALGNVILWSRPVQINASLHPEPSKINIFHRDCK 660

QY 661 RSGRDATCLAAFLCFTPIFLAPHFQTTTVGIRYNATWDERRYTPRAHLDEGGDRFTNRAV 720
 DB 661 RSGRDATCLAAFLCFTPIFLAPHFQTTTVGIRYNATWDERRYTPRAHLDEGGDRFTNRAV 720

QY 721 LLSGGQELCERINFHVLDITADYVKPVTFSVEYSLEDDPHGPMDDGPTTLRVSVFPWNG 780
 DB 721 LLSGGQELCERINFHVLDITADYVKPVTFSVEYSLEDDPHGPMDDGPTTLRVSVFPWNG 780

QY 781 CNEDEHCVPDLVLDARSDDLPTAMEYCORVLRKPAQDCSAYTLSPDITTVFIIESTRQAV 840
 DB 781 CNEDEHCVPDLVLDARSDDLPTAMEYCORVLRKPAQDCSAYTLSPDITTVFIIESTRQAV 840

QY 841 EATLENGENAYSTVLNISOSANLOFASLLQKEDSDGSIKCVNEERLQKQVCNVSYPFF 900
 DB 841 EATLENGENAYSTVLNISOSANLOFASLLQKEDSDGSIKCVNEERLQKQVCNVSYPFF 900

QY 901 RAKAKVAFRLDSEFSKSIIFLHLEIELAAGSDNSNERDSTKEDNVAPLRFHLKYEADVLT 960
 DB 901 RAKAKVAFRLDSEFSKSIIFLHLEIELAAGSDNSNERDSTKEDNVAPLRFHLKYEADVLT 960

QY 961 RSSLSHYEVKLSLSERYDIGIPFSCIPRIQNLGLFPIHGMWKITIPATISGNRL 1020
 DB 961 RSSLSHYEVKLSLSERYDIGIPFSCIPRIQNLGLFPIHGMWKITIPATISGNRL 1020

QY 1021 KLRDPLATDEANTSCNIGNSTEYRPTPVEEDLRAPQLNHSNSDVVSINCNIRLVPNQE 1080
 DB 1021 KLRDPLATDEANTSCNIGNSTEYRPTPVEEDLRAPQLNHSNSDVVSINCNIRLVPNQE 1080

QY 1081 NFHLGLNLWLSRLKALKYKSMKIMVNAALQRFHSPFIFREEDPSRQIEFISKOEDQV 1140
 DB 1081 NFHLGLNLWLSRLKALKYKSMKIMVNAALQRFHSPFIFREEDPSRQIEFISKOEDQV 1140

QY 1141 PIWIVGSTLGGLLALLVLLALRLKLGFRSARRRREGLDPTPKVLE 1188
 DB 1141 PIWIVGSTLGGLLALLVLLALRLKLGFRSARRRREGLDPTPKVLE 1188

RESULT 2
 AAU14231
 ID AAU14231 standard; protein; 1188 AA.
 XX
 AC AAU14231;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human novel protein #102.
 XX
 KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytosolic; neuroprotective; vulnary; nootropic;
 KW anticonvulsant; antiastritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200155437-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US002623.
 XX
 PR 25-JAN-2000; 2000US-00491404.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-451939/48.
 DR N-PSDB; AAS22536.
 XX
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 XX nervous system disorders, and for regenerating bone and cartilage.
 PT
 XX Example 4; Page 578-581; 894pp; English.
 PS
 XX

CC The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour in assays to determine biological activity, to raise
 CC antibodies/elicit an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention
 XX

SQ Sequence 1188 AA;

Query Match 99.6%; Score 6196; DB 4; Length 1188;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1184; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MDPLRGLVVAWALSILWPGFTDFNMTRKPRVIGPSRTAFRGYTVQOHDISGNKWLVVGA 60
 Db 1 MDPLRGLVVAWALSILWPGFTDFNMTRKPRVIGPSRTAFRGYTVQOHDISGNKWLVVGA 60

Qy 61 PLEINGYQKTDGVYKCPVHGNCTKLNIGRVTLNLSVSKDNMRGLSLATNPKNDSFLA 120
 Db 61 PLEINGYQKTDGVYKCPVHGNCTKLNIGRVTLNLSVSKDNMRGLSLATNPKNDSFLA 120

Qy 121 CSPLWSHECGSSYYTTCMCRVNSNFRFSKTVAPALQRCQTYMDIVIVLQGSNLSYWPVE 180
 Db 121 CSPLWSHECGSSYYTTCMCRVNSNFRFSKTVAPALQRCQTYMDIVIVLQGSNLSYWPVE 180

Qy 181 VQHFLINILKFFYIGPQIQGVVQYGEDVHVEHPLNDYRSVKDVEAASHIEQGGTET 240
 Db 181 VQHFLINILKFFYIGPQIQGVVQYGEDVHVEHPLNDYRSVKDVEAASHIEQGGTET 240

Qy 241 RTAFGIEFARSEAFQKGRGKAKVMIVITDGSNDSFDLEKVIQOGERDNVTRYAVAVL 300
 Db 241 RTAFGIEFARSEAFQKGRGKAKVMIVITDGSNDSFDLEKVIQOGERDNVTRYAVAVL 300

Qy 301 GYNRRGINPETFLNEIKYIASPPDDKHFFNVDEAALKDIDVALGDRIPSLGTTNKNET 360
 Db 301 GYNRRGINPETFLNEIKYIASPPDDKHFFNVDEAALKDIDVALGDRIPSLGTTNKNET 360

Qy 361 SFGLEMSQTGFSSHVVEDGVLGAVGAYDNWGNVAVLKETSAGKVIPLRESVLYKEFPBELKN 420
 Db 361 SFGLEMSQTGFSSHVVEDGVLGAVGAYDNWGNVAVLKETSAGKVIPLRESVLYKEFPBELKN 420

Qy 421 HGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVIILFTMHNRSLSITHQAMRQQQIGSYF 480
 Db 421 HGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVIILFTMHNRSLSITHQAMRQQQIGSYF 480

Qy 481 GSEITSVDIDGCVTDVLLGAPMYNEGERGKVVYVELQRNFRVYNGTLKDSHSYQNA 540
 Db 481 GSEITSVDIDGCVTDVLLGAPMYNEGERGKVVYVELQRNFRVYNGTLKDSHSYQNA 540

Qy 541 RFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFGRSILKTPKORTASLALG 600
 Db 541 RFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFGRSILKTPKORTASLALG 600

Qy 601 LQYFGCSIHGQDLNEDGLIDLAVALGNVILWSRPVVQINASLHFEPFSKINIFHRDCK 660

Db 601 LQYFGCSIHGQDLNEDGLIDLAVALGNVILWSRPVVQINASLHFEPFSKINIFHRDCK 660
 Qy 661 RSGRDATCLAAFLCTFTFLAPHFOTTTVGRYNATMDERRYTTPRAHLDEGGDRFTNRAV 720
 Db 661 RSGRDATCLAAFLCTFTFLAPHFOTTTVGRYNATMDERRYTTPRAHLDEGGDRFTNRAV 720
 Qy 721 LLSQQEICERINFHVLDTADYVKVPTFSVEYSLEDPDGHPMDLDCGWTPTLRSVSPFWNG 780
 Db 721 LLSQQEICERINFHVLDTADYVKVPTFSVEYSLEDPDGHPMDLDCGWTPTLRSVSPFWNG 780
 Qy 781 CNEDEHCVDPDLVDARSDLPTAMEYQCVLRKPAQDCSAYTILSFDTTFTIESTQRVAV 840
 Db 781 CNEDEHCVDPDLVDARSDLPTAMEYQCVLRKPAQDCSAYTILSFDTTFTIESTQRVAV 840
 Qy 841 EATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIECVNEERRLOKQVCNVSYPFF 900
 Db 841 EATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIECVNEERRLOKQVCNVSYPFF 900
 Qy 901 RAKAVAPRLDSERFSKSIFFLHLEIETLAAGSDSNERDSTKEDNVAPLRFHLKYEADVLT 960
 Db 901 RAKAVAPRLDSERFSKSIFFLHLEIETLAAGSDSNERDSTKEDNVAPLRFHLKYEADVLT 960
 Qy 961 RSSLSLHYEVKLNLSLERYDGIQPPFCIFRIQNLGLPPIHGMKMTIPIATRSNRL 1020
 Db 961 RSSLSLHYEVKLNLSLERYDGIQPPFCIFRIQNLGLPPIHGMKMTIPIATRSNRL 1020
 Qy 1021 KLRFDLTDEANTSCNIWGNSTERYPTVEEDLRAPQLNHSNDVVSINCNIRLVPNOEI 1080
 Db 1021 KLRFDLTDEANTSCNIWGNSTERYPTVEEDLRAPQLNHSNDVVSINCNIRLVPNOEI 1080
 Qy 1081 NFHLGLNLWLSLKALKYKSKIMVNAALQRFHSPIFREEDPSRQIETEPISKQEDMQV 1140
 Db 1081 NFHLGLNLWLSLKALKYKSKIMVNAALQRFHSPIFREEDPSRQIETEPISKQEDMQV 1140
 Qy 1141 PIWIIIVGSTLGGLLALLVLALRLKGLGFFRSARRRREGDPTPKVLE 1188
 Db 1141 PIWIIIVGSTLGGLLALLVLALRLKGLGFFRSARRRREGDPTPKVLE 1188

RESULT 3
 AAB50085
 ID AAB50085 standard; protein; 1188 AA.
 XX
 AC AAB50085;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE Human A259.
 XX
 KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
 KW Kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
 KW rheumatoid arthritis.
 XX
 OS Homo sapiens.

Key	Location/Qualifiers
Domain	1..1141
Peptide	/label= Extracellular_domain
Protein	1..22
Protein	/label= Signal_peptide
Protein	23..1188
Protein	/label= Mature_protein
Domain	39..74
Domain	/label= Integrin_alphasubunit_repeat_domain_#1
Domain	115..157
Domain	/label= Integrin_alphasubunit_repeat_domain_#2
Domain	164..345
Domain	/label= I_domain
Domain	367..392
Domain	/label= Integrin_alphasubunit_repeat_domain_#3
Domain	421..455
Domain	/label= Integrin_alphasubunit_repeat_domain_#4

Domain	478..516																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
--------	----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Tue Sep 21 14:48:54 2004

antidiabetic; anticonvulsant; antiparkinsonian.

Homo sapiens.

Key Location/Qualifiers

1. .1141
/note= "Extracellular domain"
1. .22
/note= "Signal peptide"
23. .1188
/note= "Mature human A259"
37. .90
states that this domain exists in human A549"
115. .157
/note= "Integrin alpha repeat domain"
164. .345
/note= "I domain or Von Willebrand Factor type A domain"
367. .392
/note= "Integrin alpha repeat domain"
421. .472
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
476. .532
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
538. .593
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
600. .654
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
1142. .1164
/note= "Transmembrane domain"
1165. .1188
/note= "Cytoplasmic domain"

WO200181414-A2.

01-NOV-2001.

27-APR-2001; 2001WO-US013516.

27-APR-2000; 2000US-00561263.

(MILL-) MILLENNIUM PHARM INC.

Pan Y, Lora J;

WPI; 2002-041397/05.

N-PSDB; AAS16873.

New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases.

Claim 9; Fig 1; 168pp; English.

The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis, particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and haematopoietic disorders (such as acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the

neural tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the human A259 polypeptide

Sequence 1188 AA;

Query Match	Score	DB	Length
Best Local Similarity	99.6%	6196;	1188;
Matches	1184;	Conservative	0; Mismatches 4; Indels 0; Gaps 0;
CC	1	MDLPRGLVVAWALS	60
CC	1	MDLPRGLVVAWALS	60
CC	61	PLETNGYQKTGDVYKCPVIHGNCTKLNLRVTL	120
CC	61	PLETNGYQKTGDVYKCPVIHGNCTKLNLRVTL	120
CC	121	CSPLWSEHCSSYYTTGMC	180
CC	121	CSPLWSEHCSSYYTTGMC	180
CC	181	VOHFLINILKXFIYIPGOIQVGVQYGEDV	240
CC	181	VOHFLINILKXFIYIPGOIQVGVQYGEDV	240
CC	241	RTAFGIEFARGEAFQKGRKGAKVMIVITD	300
CC	241	RTAFGIEFARGEAFQKGRKGAKVMIVITD	300
CC	301	GYNRRGINPETFLNEIKYIASDDPDKHFN	360
CC	301	GYNRRGINPETFLNEIKYIASDDPDKHFN	360
CC	361	SFGLMSQTFGSHVVEDGVLGAVGYDNGA	420
CC	361	SFGLMSQTFGSHVVEDGVLGAVGYDNGA	420
CC	421	HGAYLGTYTTSVSSRQGRVYVAGAPR	480
CC	421	HGAYLGTYTTSVSSRQGRVYVAGAPR	480
CC	481	GSEITSDVDIGDGVTVLLVGA	540
CC	481	GSEITSDVDIGDGVTVLLVGA	540
CC	541	RFSSSTASVREDLNQDSYNDVWCAPLED	600
CC	541	RFSSSTASVREDLNQDSYNDVWCAPLED	600
CC	601	LQYFGCSIHGQDLNEDGLIDLAVGAL	660
CC	601	LQYFGCSIHGQDLNEDGLIDLAVGAL	660
CC	661	RSGRDATCLAAFLCFTPIFLAPHQTT	720
CC	661	RSGRDATCLAAFLCFTPIFLAPHQTT	720
CC	721	LLSSGQELCERINFEHVLDTADYVKP	780
CC	721	LLSSGQELCERINFEHVLDTADYVKP	780
CC	781	CNEDEHCVDPDLVLDARS	840
CC	781	CNEDEHCVDPDLVLDARS	840
CC	841	EATLENRGENAYSTVLNISQSANLQFAS	900
CC	841	EATLENRGENAYSTVLNISQSANLQFAS	900
CC	901	RAKAKVAFRLDSEFSKSIPLHLELELA	960
CC	901	RAKAKVAFRLDSEFSKSIPLHLELELA	960

Qy	961	RSSLSHYEVKNSLERYDIGIPFSCIFRIQNLGLFPIHGMMKTIPIATRSNELL	1020
Db	961	RSSLSHYEVKNSLERYDIGIPFSCIFRIQNLGLFPIHGMMKTIPIATRSNELL	1020
Qy	1021	KLRFELTDEANTSCNIWGNSTERYPTVEEDLRAPOLNHSNSDVSINCNIRLVPNQEI	1080
Db	1021	KLRFELTDEANTSCNIWGNSTERYPTVEEDLRAPOLNHSNSDVSINCNIRLVPNQEI	1080
Qy	1081	NFHLGNLWLRSLKALKYKSMKVMNAALQRFHSPPIFREEDPSQIFEIFSKQEDWQV	1140
Db	1081	NFHLGNLWLRSLKALKYKSMKVMNAALQRFHSPPIFREEDPSQIFEIFSKQEDWQV	1140
Qy	1141	PIWIIWGSTLGGILLIALIVLARKLGFRSARRRREPGLDPTPKYLE	1188
Db	1141	PIWIIWGSTLGGILLIALIVLARKLGFRSARRRREPGLDPTPKYLE	1188

CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1189 AA;

Query Match	99.5%;	Score 6192.5;	DB 4;	Length 1189;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 1185;	Conservative	0;	Mismatches	3;
			Indels	1;
			Gaps	1;

QY	1	MOLPRGLVVAWALS	WLP	GFTD	TNNMD	TKPRV	IPG	SRTAF	PGY	TWQ	QHD	ISGN	KWLV	VGA	60
DB	1	MOLPRGLVVAWALS	WLP	GFTD	TNNMD	TKPRV	IPG	SRTAF	PGY	TWQ	QHD	ISGN	KWLV	VGA	60
QY	61	PLETTNGYOKTG	DVY	KCPV	IH	GNCT	KL	NGR	VTL	SNV	SER	KDN	MLG	L	120
DB	61	PLETTNGYOKTG	DVY	KCPV	IH	GNCT	KL	NGR	VTL	SNV	SER	KDN	MLG	L	120
QY	121	CSPLWSHC	SGS	YYT	TG	MC	SRV	N	FR	SK	TVA	PAL	QRC	QY	180
DB	121	CSPLWSHC	SGS	YYT	TG	MC	SRV	N	FR	SK	TVA	PAL	QRC	QY	180
QY	181	VOHFLIN	L	K	PY	I	G	P	G	Q	I	Q	V	G	240
DB	181	VOHFLIN	L	K	PY	I	G	P	G	Q	I	Q	V	G	240
QY	241	RTAFGIE	P	AR	SE	A	F	K	G	R	K	A	K	W	300
DB	241	RTAFGIE	P	AR	SE	A	F	K	G	R	K	A	K	W	300
QY	301	GYNNR	G	I	N	P	E	T	F	L	E	I	K	Y	360
DB	301	GYNNR	G	I	N	P	E	T	F	L	E	I	K	Y	360
QY	361	SFGL	E	M	S	Q	T	G	F	S	H	V	E	D	420
DB	361	SFGL	E	M	S	Q	T	G	F	S	H	V	E	D	420
QY	421	HGAY	L	G	Y	T	V	T	S	V	S	S	R	G	480
DB	421	HGAY	L	G	Y	T	V	T	S	V	S	S	R	G	480
QY	481	GS	Y	T	S	V	D	I	D	G	V	T	D	V	540
DB	481	GS	Y	T	S	V	D	I	D	G	V	T	D	V	540
QY	541	RF	G	S	S	I	A	S	V	R	D	L	N	O	600
DB	541	RF	G	S	S	I	A	S	V	R	D	L	N	O	600
QY	601	LQ	V	F	G	S	I	E	G	O	L	D	N	E	660
DB	601	LQ	V	F	G	S	I	E	G	O	L	D	N	E	660
QY	661	R	S	G	R	A	T	C	L	A	A	F	C	T	720
DB	661	R	S	G	R	A	T	C	L	A	A	F	C	T	720
QY	721	L	S	S	G	O	E	L	C	E	R	I	N	F	780
DB	721	L	S	S	G	O	E	L	C	E	R	I	N	F	780
QY	781	C	N	E	D	E	H	C	V	P	D	L	V	L	840
DB	781	C	N	E	D	E	H	C	V	P	D	L	V	L	840
QY	841	E	A	T	L	E	N	G	E	N	A	I	S	T	900
DB	841	E	A	T	L	E	N	G	E	N	A	I	S	T	900
QY	901	R	A	K	A	V	A	P	R	D	S	E	F	S	960
DB	901	R	A	K	A	V	A	P	R	D	S	E	F	S	960

Db 901 RAKAKVAFRLDPFSKIFLHLLLEIAAGSDSNERDSTKEDNVAPLRFHLKYEADVLFT 960
Qy 961 RSSLSHYEVKLNSSLRYDYGIPFSCIFRIQNLGLFPIHGMMKTTIPIATRSNGNRL 1020
Db 961 RSSLSHYEVKLNSSLRYDYGIPFSCIFRIQNLGLFPIHGMMKTTIPIATRSNGNRL 1020
Qy 1021 KLRFPLTDE-ANTSNIWGNSTYRPTPEVEDLRRAPOLNHSNDVVSINCLRLVPOE 1079
Db 1021 KLRFPLTDE-ANTSNIWGNSTYRPTPEVEDLRRAPOLNHSNDVVSINCLRLVPOE 1080
Qy 1080 INFHLGLNMLRSKALKYSKMKIMVNAALQROPHSPFIFREEDPSRQIFEIFISKQEDWQ 1139
Db 1081 INFHLGLNMLRSKALKYSKMKIMVNAALQROPHSPFIFREEDPSRQIFEIFISKQEDWQ 1140
Qy 1140 VPIWIVGSTLGGLLLLALLVIALRKUGFRRSARRREPGLDPTPKVLE 1188
Db 1141 VPIWIVGSTLGGLLLLALLVIALWKLGFRSARRREPGLDPTPKVLE 1189

RESULT 6

AAU14467
ID AAU14467 standard; protein; 1188 AA.

AC AAU14467;

DT 24-OCT-2001 (first entry)

DE Human novel protein #338.

KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
Parkinson's disease; immunogen; antibody; gene therapy; neurological disorder;
tissue regeneration; immune disorder.

OS Homo sapiens.

PN WO200155437-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US002623.

PR 25-JAN-2000; 2000US-00491404.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-451939/48.

DR N-PSDB; AAS22772.

PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage.

PS Example 4; Page 828-831; 894pp; English.

CC The invention relates to polynucleotides encoding novel human proteins or
CC their active domains. The polypeptides, polynucleotides and antibodies
CC raised against the polypeptides are used in a method of treatment of a
CC mammal and prevention of disorders caused by the aberrant protein
CC expression or activity. The polypeptides can be used as molecular weight
CC markers, food supplements, and in antibody production. The polypeptides
CC are used to identify compounds which bind to the polypeptides.
CC Polynucleotides of the invention are used as probes and primers, for
CC sequencing, for chromosome or gene mapping, in the production of
CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
CC therapy. Polypeptides of the invention can be used to target drugs to a
CC tumour, in assays to determine biological activity, to raise
CC antibodies/elicite an immune response, to determine quantitative protein
CC levels, as tissue markers, and to isolate receptors or ligands.

CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
CC diseases, nervous system disorders, and infection. The present sequence
CC represents a protein of the invention

XX SQ Sequence 1188 AA;

Query Match 99.5%; Score 6192; DB 4; Length 1188;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1183; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MDLPRGLVAVWALSLLWPGFTDTFNMTRKPRVPGSRTAFFGYTVQGHDSGNKWLVVGA 60

Db 1 MDLPRGLVAVWALSLLWPGFTDTFNMTRKPRVPGSRTAFFGYTVQGHDSGNKWLVVGA 60

Qy 61 PLETNGYQKTGDVYKCPVIHGNCTKLNIGRVTLNSVSRKDNMRLGLSLATNPKNDSFLA 120

Db 61 PLETNGYQKTGDVYKCPVIHGNCTKLNIGRVTLNSVSRKDNMRLGLSLATNPKNDSFLA 120

Qy 121 CSPLWSHECGSSYYTTCMSRVNSNFRFSKTVAPALORCOTYMDIVIVLDSNSIYPWVE 180

Db 121 CSPLWSHECGSSYYTTCMSRVNSNFRFSKTVAPALORCOTYMDIVIVLDSNSIYPWVE 180

Qy 181 VOHFLIMILKFFYIGPGQIOGVVQYGEDVVEPHLNDYRSVKDVEAAASHIEORGGET 240

Db 181 VOHFLIMILKFFYIGPGQIOGVVQYGEDVVEPHLNDYRSVKDVEAAASHIEORGGET 240

Qy 241 RTAFGIEFARSEAFQKGRGAKKVMIVITDGEHSDSPDLKVIQQSERDNTVRYAVAVL 300

Db 241 RTAFGIEFARSEAFQKGRGAKKVMIVITDGEHSDSPDLKVIQQSERDNTVRYAVAVL 300

Qy 301 GYNNRGINPETFLNEIKYIASDPDDKHFFNVTDAAALKDIVDALGDRIFSLGTTNKNET 360

Db 301 GYNNRGINPETFLNEIKYIASDPDDKHFFNVTDAAALKDIVDALGDRIFSLGTTNKNET 360

Qy 361 SFGLEMSQTGFSSHWEDGVLGAVGYDMNGAVLKEKTSAGKVIPLRESYLKEPPEELKN 420

Db 361 SFGLEMSQTGFSSHWEDGVLGAVGYDMNGAVLKEKTSAGKVIPLRESYLKEPPEELKN 420

Qy 421 HGAYLGYTVTSVSSRQGRVYVAGAPRNHTGKVLFTMHNNRSLTIHQAMRGQOIGSYF 480

Db 421 HGAYLGYTVTSVSSRQGRVYVAGAPRNHTGKVLFTMHNNRSLTIHQAMRGQOIGSYF 480

Qy 481 GSEITSVDIDGCVTDVLLVCAPMYFNEGRERGVYVYELQNRFRVYNGTLKDSHSYQNA 540

Db 481 GSEITSVDIDGCVTDVLLVCAPMYFNEGRERGVYVYELQNRFRVYNGTLKDSHSYQNA 540

Qy 541 RFGSISAVRDLNQSNDVNVVGGAPLEDNHAAGIYIFHGFRGSILKTPKQITASELATG 600

Db 541 RFGSISAVRDLNQSNDVNVVGGAPLEDNHAAGIYIFHGFRGSILKTPKQITASELATG 600

Qy 601 LQYFGCSIHGQLDNLNEDGLIDLAVGALGNVILMSRPVQVNASLHFPFSKINIFHRDCK 660

Db 601 LQYFGCSIHGQLDNLNEDGLIDLAVGALGNVILMSRPVQVNASLHFPFSKINIFHRDCK 660

Qy 661 RSGRDATCLAAFLCFTPIFLAPHQTTTVGTRYNATMDERYTPRAHLDGEGDRTNRAV 720

Db 661 RSGRDATCLAAFLCFTPIFLAPHQTTTVGTRYNATMDERYTPRAHLDGEGDRTNRAV 720

Qy 721 LLSQQEUCERINPHVLDITADYVVKPVPFSVEYSLDDPHGPMDDGWPPTTLRVSPVFWNG 780

Db 721 LLSQQEUCERINPHVLDITADYVVKPVPFSVEYSLDDPHGPMDDGWPPTTLRVSPVFWNG 780

Qy 781 CNEDEHCVPLDLVARSDLPTAMEYQCVRLRKPAQDCSAYTSLSDFTTVFIESTQRVAV 840

Db 781 CNEDEHCVPLDLVARSDLPTAMEYQCVRLRKPAQDCSAYTSLSDFTTVFIESTQRVAV 840

QY 841 EATLENGENAYSTVLNISOSANLQFASLIQKEDSDGSECVNEERLQKQCNVSPFF 900
 DB 841 EATLENGENAYSTVLNISOSANLQFASLIQKEDSDGSECVNEERLQKQCNVSPFF 900
 QY 901 RAKAKVAFRLDSEFFSKIFLHLEIELAAGSDSNERDSTKEDNVAPLRFHLKYEADVLEF 960
 DB 901 RAKAKVAFRLDSEFFSKIFLHLEIELAAGSDSNERDSTKEDNVAPLRFHLKYEADVLEF 960
 QY 961 RSSLSHYEVKLNSSIERVDGIGPPFSCIFRIQNLGLFPIHGMMKITTIPIATRSNRL 1020
 DB 961 RSSLSHYEVKLNSSIERVDGIGPPFSCIFRIQNLGLFPIHGMMKITTIPIATRSNRL 1020
 QY 1021 KLKDFLTDEANTSCNIWGNSTERYPTVEEDLRAPQLNHSNDDVVSINCNIRLVNQE 1080
 DB 1021 KLKDFLTDEANTSCNIWGNSTERYPTVEEDLRAPQLNHSNDDVVSINCNIRLVNQE 1080
 QY 1081 NFHLNLNLRSLKALKYKSMKIMVNAALQRFHSPFIFREEDPSRQIVFEISKQEDQV 1140
 DB 1081 NFHLNLNLRSLKALKYKSMKIMVNAALQRFHSPFIFREEDPSRQIVFEISKQEDQV 1140
 QY 1141 PIWIIVGSTLGGLLALLVLALRKLGFFRSARRRREPGLDTPPKYLE 1188
 DB 1141 PIWIIVGSTLGGLLALLVLALRKLGFFRSARRRREPGLDTPPKYLE 1188

RESULT 7 ADE09956

ID ADE09956 standard; protein; 1188 AA.

XX ADE09956;

XX 29-JAN-2004 (first entry)

XX Novel protein-related contig polypeptide sequence #544.

XX novel gene; novel protein; tissue marker; molecular weight marker;

XX chromosome marker; genetic disorder; contig.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

XX 14-MAR-2002; 2002US-0365091P.

XX 14-MAR-2002; 2002US-0365384P.

XX 12-APR-2002; 2002US-0372381P.

XX 12-APR-2002; 2002US-0372615P.

XX 22-APR-2002; 2002US-00128558.

XX 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.

XX Disclosure; SEQ ID NO 3022; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially

CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence was used in the
 CC exemplification of the invention.

SQ Sequence 1188 AA;

Query Match 99.5%; Score 6192; DB 7; Length 1188;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1183; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MDLPGLVAVAWALSHPGETDFNMDTRKPRVIPSRTAFFGTYTVOQHDISGNKWLAVGA 60

DB 1 MDLPGLVAVAWALSHPGETDFNMDTRKPRVIPSRTAFFGTYTVOQHDISGNKWLAVGA 60

QY 61 PLETTYGYQKTGVYKCPVIHGNCTKLNLRVTLNVSEKDNMRGLSLATNPKNSFLA 120

DB 61 PLETTYGYQKTGVYKCPVIHGNCTKLNLRVTLNVSEKDNMRGLSLATNPKNSFLA 120

QY 121 CSPLMSHECGSSYYTTGMCNRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180

DB 121 CSPLMSHECGSSYYTTGMCNRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180

QY 181 VOHELINILKKFYIGPGQIOGVVOYGEDVHVEFHLNDYRSYKDVVEAASHIEQCGTET 240

DB 181 VOHELINILKKFYIGPGQIOGVVOYGEDVHVEFHLNDYRSYKDVVEAASHIEQCGTET 240

QY 241 RTAFGIEFARSEAFQGGKGAKKVMIVITDGESHSDPLEKVIQOESRDNVTRAVAVL 300

DB 241 RTAFGIEFARSEAFQGGKGAKKVMIVITDGESHSDPLEKVIQOESRDNVTRAVAVL 300

QY 301 GYNNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDIVDALGDRIFSLGNTKNET 360

DB 301 GYNNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDIVDALGDRIFSLGNTKNET 360

QY 361 SFGLEMSQTGFSSHHVVEDGVLGAGAYDNGAVLKETSAGKVIPLRESYLKEFPEELKN 420

DB 361 SFGLEMSQTGFSSHHVVEDGVLGAGAYDNGAVLKETSAGKVIPLRESYLKEFPEELKN 420

QY 421 HGAYLGYTVTSVSSRQGRVYVAGAPRFNHTGKVLFTMNNRSLTIHQAMEGQOIGSYF 480

DB 421 HGAYLGYTVTSVSSRQGRVYVAGAPRFNHTGKVLFTMNNRSLTIHQAMEGQOIGSYF 480

QY 481 GSEITSDIDGEGVTDVLLVGAPVFNREGRGKVVYVELQNRVFNVTGLKDSHSYQNA 540

DB 481 GSEITSDIDGEGVTDVLLVGAPVFNREGRGKVVYVELQNRVFNVTGLKDSHSYQNA 540

QY 541 RFGSSIASVRDLNODSYNDVVVGAPLEDNHAGAIYIFHGFGSILKTPKORTASELATG 600

DB 541 RFGSSIASVRDLNODSYNDVVVGAPLEDNHAGAIYIFHGFGSILKTPKORTASELATG 600

QY 601 LQYFGCSIHGQLDNEDGLIDLAVGALGNVILMSRPVYVQINASLHFEPFSKINIPHRDCK 660

DB 601 LQYFGCSIHGQLDNEDGLIDLAVGALGNVILMSRPVYVQINASLHFEPFSKINIPHRDCK 660

QY 661 RSGGDTCLAAFLCFTPTIFLAPHFQTTVGRYRNATMDERRYTPRAHLDGEGDRPTNRAV 720

DB 661 RSGGDTCLAAFLCFTPTIFLAPHFQTTVGRYRNATMDERRYTPRAHLDGEGDRPTNRAV 720

QY 721 LLSSGQELCERINFHVLDTADYVVKPVTFVSVEYSLEDDPHGPMLDGWTTLRVSVFPWNG 780

DB 721 LLSSGQELCERINFHVLDTADYVVKPVTFVSVEYSLEDDPHGPMLDGWTTLRVSVFPWNG 780

QY 781 CNEDEHCVPLDVLDAARSDLPAMEYCCQRLKPAQDCSAYTSLFDTTTFITESTQRVAV 840

DB 781 CNEDEHCVPLDVLDAARSDLPAMEYCCQRLKPAQDCSAYTSLFDTTTFITESTQRVAV 840

QY 841 EATLENGENAYSTVLNISOSANLQFASLIQKEDSDGSECVNEERLQKQCNVSPFF 900

DB 841 EATLENGENAYSTVLNISOSANLQFASLIQKEDSDGSECVNEERLQKQCNVSPFF 900

QY 901 RAKAKVAFRLDSEFFSKIFLHLEIELAAGSDSNERDSTKEDNVAPLRFHLKYEADVLEF 960

Db 901 RAKAKAFRLDFEFSKIFLHLLEIELAAGSDSNERUSTKEDNVAPLRFHLKHYEVDVLT 960
Qy 961 RSSLSHYEVLKNSLERYDGGPPFCIFRIQNLGLFPIHGMWMTIPIATRSNRL 1020
Db 961 RSSLSHYEVLKNSLERYDGGPPFCIFRIQNLGLFPIHGMWMTIPIATRSNRL 1020
Qy 1021 KLRFELDEANTSCNIWGNSTYRPTVEEDLRAPQLNHSNDVVSINCLRLVPNQEI 1080
Db 1021 KLRFELDEANTSCNIWGNSTYRPTVEEDLRAPQLNHSNDVVSINCLRLVPNQEI 1080
Qy 1081 NFHLGLNLWLSLKALKYKSMKWNAALQRFSPFIREFDPSRQIPEISKQEDWQV 1140
Db 1081 NFHLGLNLWLSLKALKYKSMKWNAALQRFSPFIREFDPSRQIPEISKQEDWQV 1140
Qy 1141 PIWIVGSTLGLLLALLLVLLALRLKGFRRARRRREPGLDPTPKVLE 1188
Db 1141 PIWIVGSTLGLLLALLLVLLALRLKGFRRARRRREPGLDPTPKVLE 1188

RESULT 8
ID AAB25582
XX AAB25582 standard; protein; 1189 AA.
AC
XX
DT 21-NOV-2000 (first entry)
XX

ITGall protein encoded by human secreted protein gene #7.

Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
anti-rheumatic; dermatological; antiproliferative; antiarteriosclerotic;
anticancer; vulnery; antiviral; antibacterial; antifungal;
immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
Crohn's disease; nephritis; hyperproliferative disorder;
cardiovascular disorder; coronary arteriosclerosis; myocardiitis; cancer;
melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
Homo sapiens.

OS
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX 27-OCT-1999; 99WO-US025031.
XX
XX 28-OCT-1998; 98US-0105971P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.

PA
XX
XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
PI Greene JM;

XX
XX WPI; 2000-387742/33.
XX
XX N-PSDB; AAA80612.

XX
XX Isolated nucleic acid molecules encoding human secreted proteins are used
PT for the prevention, amelioration and treatment of autoimmune,
PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
PT wounds, and infectious diseases.

XX
XX Claim 1; Fig 19A-F; 803pp; English.

XX
XX The present invention relates to 12 secreted human proteins and the
CC nucleotide sequences encoding them. The polynucleotide sequences given in
CC AAA80606-A80623 encode the 12 secreted protein sequences given in
CC AAB25576-B25593. The human secreted proteins have various activities
CC dependent on the tissues in which they are expressed. Examples of the
CC activities of the proteins include: immunosuppressant; anti-inflammatory;
CC antiarthritic; anti-rheumatic; dermatological; antiproliferative;
CC antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial;
CC and antifungal activity. The proteins, polypeptides, agonists and

CC antagonists may be used to treat prevent and/or diagnose various disease,
CC disorders and conditions examples of which include: immune disorders e.g.
CC Addison's disease, rheumatoid arthritis, dermatitis, and multiple
CC sclerosis; inflammatory disorders e.g. inflammatory bowel disease,
CC Crohn's disease and nephritis; hyperproliferative disorders such as
CC paraproteinemia and purpura; cardiovascular disorders e.g. coronary
CC arteriosclerosis and myocardiitis; cancer e.g. melanoma and lymphoma. The
CC proteins and polynucleotide sequences may also be used in wound healing
CC and the treatment of infectious diseases. The human secreted protein gene
CC #7 and protein sequences are represented in sequences AAA80612 and
CC AAB25582. Secreted protein gene #7 is located at position chromosome 15
CC q22.3-23. Sequences AAA80652-A80661 represent genes which are related to
CC the secreted protein gene#7
XX
XX Sequence 1189 AA;

Query Match 99.4%; Score 6188.5; DB 3; Length 1189;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MDLPRGLVWAWLSLWPGFTDTFNMOTRKPRVIFGSRTAFFGYTVQOHDISGNKWLWVGA 60
Db 1 MDLPRGLVWAWLSLWPGFTDTFNMOTRKPRVIFGSRTAFFGYTVQOHDISGNKWLWVGA 60
Qy 61 PLETNQYKQTDGVYKCPVIHGNCTKLNLRVTLSNVSERKDNMRGLSLATNPKNDSFLA 120
Db 61 PLETNQYKQTDGVYKCPVIHGNCTKLNLRVTLSNVSERKDNMRGLSLATNPKNDSFLA 120
Qy 121 CSPLWSHECGSSYYTTGMSRVNSNFRPSKTVAPALQRCQTYMDIVLDGNSNIYPWVE 180
Db 121 CSPLWSHECGSSYYTTGMSRVNSNFRPSKTVAPALQRCQTYMDIVLDGNSNIYPWVE 180
Qy 181 VOHFLINILKXFYIGPGQIQGVVQYQGVDDVHVEHLNDYRSVKDVVEAASHIEQGGTET 240
Db 181 VOHFLINILKXFYIGPGQIQGVVQYQGVDDVHVEHLNDYRSVKDVVEAASHIEQGGTET 240
Qy 241 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGHSDSPDLKVIQOQSRDNNVTRYAVVL 300
Db 241 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGHSDSPDLKVIQOQSRDNNVTRYAVVL 300
Qy 301 GYNNRRGINPETFLNEIKYIASDPPDKHFFNVTDAAALKDIVDALGDRIFSLGNTKNET 360
Db 301 GYNNRRGINPETFLNEIKYIASDPPDKHFFNVTDAAALKDIVDALGDRIFSLGNTKNET 360
Qy 361 SFGLEMSQTGFSSHVVEDGLVGAVYDNGAVLKETSAKVPLRESYLKEPPEELKN 420
Db 361 SFGLEMSQTGFSSHVVEDGLVGAVYDNGAVLKETSAKVPLRESYLKEPPEELKN 420
Qy 421 HGAYLGYTVTSVSSRQGRVYVAGAPRENHTGKVILFTMNNRSLTIHQAMRGQIGSYF 480
Db 421 HGAYLGYTVTSVSSRQGRVYVAGAPRENHTGKVILFTMNNRSLTIHQAMRGQIGSYF 480
Qy 481 GSEITSVDIDGQVTDVLLVGAPMYFNEGRGKGVYVELQNRFPVYNGTLDKSHSYQNA 540
Db 481 GSEITSVDIDGQVTDVLLVGAPMYFNEGRGKGVYVELQNRFPVYNGTLDKSHSYQNA 540
Qy 541 RFGSSIASVRDLNODSYNDVVGAPLEDNAGAVIYIHPGERSILKTPKORTASLSELATG 600
Db 541 RFGSSIASVRDLNODSYNDVVGAPLEDNAGAVIYIHPGERSILKTPKORTASLSELATG 600
Qy 601 LQYFGCSIHGQDLNEDGLIDLAVGALGNVILMSRPVQVNASLHFFPSKINIFHRDCK 660
Db 601 LQYFGCSIHGQDLNEDGLIDLAVGALGNVILMSRPVQVNASLHFFPSKINIFHRDCK 660
Qy 661 RSGRDATCLAAFLCPTPIFLAPHFOTTTVGIRYNATMDERRYTPRAHIDEGGDRFTNRAV 720
Db 661 RSGRDATCLAAFLCPTPIFLAPHFOTTTVGIRYNATMDERRYTPRAHIDEGGDRFTNRAV 720
Qy 721 LLSSQGEICERINFHVLDTADYVKVTSVESLDDPDHGMDDGPTTLRVSPFWNG 780
Db 721 LLSSQGEICERINFHVLDTADYVKVTSVESLDDPDHGMDDGPTTLRVSPFWNG 780
Qy 781 CNEDEHCVPDVLVDARSDLPTAMEYCORVLRKPAQDCSAYTILSFDTTVTFTIESTQRVAV 840


```
Db 781 CNEDEHCVDFVLDAESDLFTAMEYQORVLRPAQDCSAYTSFDTVFIIESTRORAV 840
QY 841 EATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIQECVNEERLQKQVCNVSYPFF 900
Db 841 EATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIQECVNEERLQKQVCNVSYPFF 900
QY 901 RAKAKVAFRLDFEFSKSIPLHLEIEIELAAGSDSNERDSTKEDNVAFLRHLKYEADVLFT 960
Db 901 RAKAKVAFRLDFEFSKSIPLHLEIEIELAAGSDSNERDSTKEDNVAFLRHLKYEADVLFT 960
QY 961 RGSLSHYEVKLNSSLYERVDGIPGPPSCIFRIQNLGLFPIHGMWKITIPATRSNRL 1020
Db 961 RGSLSHYEVKLNSSLYERVDGIPGPPSCIFRIQNLGLFPIHGMWKITIPATRSNRL 1020
QY 1021 KLKDELTDLDE-ANTSCNIWGNSTYRTPVVEEDLRRAPQLNHSNDSVVSINCLRLVNOE 1079
Db 1021 KLKDELTDLDEVANTSCNIWGNSTYRTPVVEEDLRRAPQLNHSNDSVVSINCLRLVNOE 1080
QY 1080 INFHLLGNLWLSRLSKALKYKSMKIMVNAALQRFHSPFIREDPSRQIEFISKOEDWQ 1139
Db 1081 INFHLLGNLWLSRLSKALKYKSMKIMVNAALQRFHSPFIREDPSRQIEFISKOEDWQ 1140
QY 1140 VPIWIIVGSTGLGLLLALLVLALRLKLGFFRSARRRREPGLDPTPKVLE 1188
Db 1141 VPIWIIVGSTGLGLLLALLVLALRLKLGFFRSARRRREPGLDPTPKVLE 1189

RESULT 9
ABR58364
ID ABR58364 standard; protein; 1189 AA.
AC ABR58364;
XX
XX
DT 07-JUL-2003 (first entry)
DE Human NOV2a.
KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;
KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;
KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;
KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;
KW neurodegenerative disorder; Alzheimer's disease; immune disorder;
KW haematopoietic disorder.
XX
OS Homo sapiens.
XX
XX WO2003029423-A2.
PN
PD 10-APR-2003.
XX
XX 02-OCT-2002; 2002WO-US031358.
PF
XX
PR 02-OCT-2001; 2001US-0326483P.
PR 05-OCT-2001; 2001US-0327342P.
PR 09-OCT-2001; 2001US-0327917P.
PR 09-OCT-2001; 2001US-0328029P.
PR 09-OCT-2001; 2001US-0328044P.
PR 09-OCT-2001; 2001US-0328056P.
PR 12-OCT-2001; 2001US-0328849P.
PR 15-OCT-2001; 2001US-0329414P.
PR 17-OCT-2001; 2001US-0330142P.
PR 22-OCT-2001; 2001US-0341058P.
PR 24-OCT-2001; 2001US-0339266P.
PR 24-OCT-2001; 2001US-0343629P.
PR 29-OCT-2001; 2001US-0349575P.
PR 01-NOV-2001; 2001US-0346357P.
PR 12-APR-2002; 2002US-0371972P.
PR 12-APR-2002; 2002US-0371980P.
PR 17-APR-2002; 2002US-0373261P.
PR 19-APR-2002; 2002US-0373805P.
PR 23-APR-2002; 2002US-0374738P.
PR 16-MAY-2002; 2002US-0381101P.
```

```
PR 17-MAY-2002; 2002US-0381635P.
PR 29-MAY-2002; 2002US-0383830P.
PR 01-OCT-2002; 2002US-00262839.
PA (CURA-) CURAGEN CORP.
XX
XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
PI Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;
XX
XX WPI; 2003-381625/36.
DR N-PSDB; ACC72076.
XX
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or
PT dyslipidemia, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
PS Claim 1; Page 105; 487pp; English.
XX
CC The present invention relates to novel human NOV proteins and their
CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV
CC proteins are useful in manufacturing a medicament for treating a syndrome
CC associated with a human disease. The NOV proteins and coding sequences
CC may be used to diagnose, treat or prevent metabolic disorders such as
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative
CC disorders such as Alzheimer's disease or Parkinson's disease, immune
CC disorders, haematopoietic disorders and various dyslipidaemias
XX
XX Sequence 1189 AA;
```

```
Query Match 99.4%; Score 6188.5; DB 6; Length 1189;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDLPRGLVAVALSLLWPGFTDTFNMDTRKPRVPGSRTPAFGTYTQQDHDISGNKMLVGA 60
Db 1 MDLPRGLVAVALSLLWPGFTDTFNMDTRKPRVPGSRTPAFGTYTQQDHDISGNKMLVGA 60
QY 61 PLEITNGYQKTGDVYKCPVIHGNCTKLNLRVTLNSVSRKONMRGLSLATNPKNSFLA 120
Db 61 PLEITNGYQKTGDVYKCPVIHGNCTKLNLRVTLNSVSRKONMRGLSLATNPKNSFLA 120
QY 121 CSPLSWHECGSSYYTTGMCSSRVNSNFRFSKTVPALQRCQTYMDIVILDGNSIYPWVE 180
Db 121 CSPLSWHECGSSYYTTGMCSSRVNSNFRFSKTVPALQRCQTYMDIVILDGNSIYPWVE 180
QY 181 VQHELINILKKFYIGPGQIQGVVQYGEDVHFEHLNDYRSVKDVVEAASHIEQRGTTET 240
Db 181 VQHELINILKKFYIGPGQIQGVVQYGEDVHFEHLNDYRSVKDVVEAASHIEQRGTTET 240
QY 241 RTAFGIEFARSEAFQKGGKGAKKVMIVITDGESHSDSPLEKVIQCSRDNVTRAVAVL 300
Db 241 RTAFGIEFARSEAFQKGGKGAKKVMIVITDGESHSDSPLEKVIQCSRDNVTRAVAVL 300
QY 301 GYNNRRGINPETFLNEIKYIASDDPDKHFFNVVTEAALKDIVDALGDRIFSLGNTKNET 360
Db 301 GYNNRRGINPETFLNEIKYIASDDPDKHFFNVVTEAALKDIVDALGDRIFSLGNTKNET 360
QY 361 SFGLEMSOTGFSSHVVEDGVLGAVGAYDMNGAVLKETSGAGVPLRPSYLKEPPEELKN 420
Db 361 SFGLEMSOTGFSSHVVEDGVLGAVGAYDMNGAVLKETSGAGVPLRPSYLKEPPEELKN 420
QY 421 HGAYLGYTVTSVSSRQGRVYVAGAPRNHTGKVLFTMNNRSITLHQAMRGQOIGSYF 480
Db 421 HGAYLGYTVTSVSSRQGRVYVAGAPRNHTGKVLFTMNNRSITLHQAMRGQOIGSYF 480
QY 481 GSEITSVDIDGGVTDVLLVGAPMYFNREGRGKVVYVELQNRFPVYNGTLKDSHYQNA 540
Db 481 GSEITSVDIDGGVTDVLLVGAPMYFNREGRGKVVYVELQNRFPVYNGTLKDSHYQNA 540
```



```
QY 541 RFSSIASVRDLNODSYNDVVVVGAPLEDNHAGAIYIFHGFRGSILTKPKORITASELATG 600
Db |||
QY 541 RFSSIASVRDLNODSYNDVVVVGAPLEDNHAGAIYIFHGFRGSILTKPKORITASELATG 600
Db |||
QY 601 LQYFGCSIHGQLDLNEDGLIDLAVGALGNVILMSRPVQVNASLHFFSPKINIFHRDCK 660
Db |||
QY 601 LQYFGCSIHGQLDLNEDGLIDLAVGALGNVILMSRPVQVNASLHFFSPKINIFHRDCK 660
Db |||
QY 661 RSGRDATCLAAFLCFTPTIFLAPHFQTTTIGRYNATMDERRYTPRAHLDEGGDRFTNRAV 720
Db |||
QY 721 LLSSGOELCERINHFVLDTADYVKVPVTSVEYSLEDPDHGMMLDDGWTTLRVSPVFWNG 780
Db |||
QY 781 CNEDEHCVPLDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTSLPDTTFFIESTRQVAV 840
Db |||
QY 841 EATLENGENAYSTVLNISQSANLOFASLIOKEDSDGSIECVNEERRLQKQVCNVSPYFF 900
Db |||
QY 841 EATLENGENAYSTVLNISQSANLOFASLIOKEDSDGSIECVNEERRLQKQVCNVSPYFF 900
Db |||
QY 901 RAKAVAPRLDSFSSKIFLHLELELAAGSDSNERDSTKEDNVAPLRFHLKYRADVLFT 960
Db |||
QY 901 RAKAVAPRLDSFSSKIFLHLELELAAGSDSNERDSTKEDNVAPLRFHLKYRADVLFT 960
Db |||
QY 961 RSSLSHYEVKLSLEYDGIQPPFCIFRIQNLGLPFIHGMMKTIPIATRSNGRLL 1020
Db |||
QY 1021 KLRDLTDE-ANTSCNINWSTETPTVEEDLRAPQLNHSNVVWVINCINIRLVNQE 1079
Db |||
QY 1021 KLRDLTDEVANTSCNINWSTETPTVEEDLRAPQLNHSNVVWVINCINIRLVNQE 1080
Db |||
QY 1080 INFHLLGNLWLSLKALKYKSKIMVNAALORQFHSPIFREEDPSRQIEFISKOEDWQ 1139
Db |||
QY 1081 INFHLLGNLWLSLKALKYKSKIMVNAALORQFHSPIFREEDPSRQIEFISKOEDWQ 1140
Db |||
QY 1140 VPIWIVGSTLGLLLALLLALRLKGLGFFRSARRRREPGLDPTPKVLE 1188
Db |||
QY 1141 VPIWIVGSTLGLLLALLLALRLKGLGFFRSARRRREPGLDPTPKVLE 1189
Db |||

RESULT 10
ADA27054
ID ADA27054 standard; protein; 1189 AA.
XX
AC ADA27054;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human novel secreted protein from cDNA HOHB69 #1.
XX
KW Cytostatic; antiinflammatory; immunomodulator; neuroprotective;
KW hemostatic; gene therapy; cancer; inflammation; immune disorder;
KW neurological disorder; blood clotting disorder; food additive;
KW preservative; human; secreted protein.
XX
OS Homo sapiens.
XX
FN US2003055231-A1.
XX
PD 20-MAR-2003.
XX
PF 29-OCT-2001; 2001US-00984130.
XX
PR 28-OCT-1998; 98US-0105971P.
PR 27-OCT-1999; 99WO-US025031.
PR 19-APR-2000; 2000US-0198407P.
PR 30-OCT-2000; 2000US-0243792P.
PR 18-APR-2001; 2001US-00836353.
```

```
XX
PA (NIJ/) NI J.
PA (YOUN/) YOUNG P E.
PA (KENN/) KENNY J J.
PA (OLSE/) OLSEN H S.
PA (MOOR/) MOORE P A.
PA (WEIY/) WEI Y.
PA (GREE/) GREENE J M.
PA (RUBE/) RUBEN S M.
PA (LIUD/) LIU D.
PA (CROC/) CROCKER P R.
```

Ni J, Young PE, Kenny JJ, Olsen HS, Moore PA, Wei Y, Greene JM;
Ruben SM, Liu D, Crocker PR;

WPI; 2003-567103/53.
N-PSDB; ADA27036.

New human secreted nucleic acid molecules and polypeptides, useful for preventing, treating, or ameliorating a medical condition, such as cancer, inflammation, immune disorders, neurological and blood clotting disorders.

Claim 11; Fig 19; 454pp; English.

The invention relates to an isolated nucleic molecule that is at least 95% identical to 18 human cDNA sequences representing 12 novel genes encoding secreted proteins or a polynucleotide fragment of the cDNA sequence contained in American Type Culture Collection (ATCC) deposit No. defined in the specification, its species homologue, a variant or allelic variant of the polynucleotide having a polynucleotide capable of hybridising under conditions the polynucleotide, where the polynucleotide does not hybridise under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A or T residues. Also included are recombinant vectors, host cells (for producing the polypeptide), the secreted polypeptide (comprising a sequence that is at least 95% identical to a polypeptide fragment, domain, epitope, full-length protein, variant, allelic variant or species homologue), antibodies that specifically bind to the polypeptides, diagnosing, treating, preventing or ameliorating a medical condition by administering the polynucleotide or the polypeptide, the gene corresponding to the cDNA sequence and identifying an activity in a biological assay (by expressing the cDNA sequence in a cell, isolating the supernatant, and detecting an activity in a biological assay and identifying the protein in the supernatant having the activity). The polypeptides, nucleic acids and antibodies are useful for diagnosing a pathological condition or a susceptibility to a pathological condition, for preventing, treating, or ameliorating a medical condition, such as cancer, inflammation and other immune disorders, neurological and blood clotting disorders (many examples are given in the specification). The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues immunohistochemistry assays. The polypeptide, polynucleotide, agonist or antagonist may also be used as a food additive or preservative to increase or decrease storage capabilities, fat content or other nutritional components. The present is a secreted protein of the invention.

Sequence 1189 AA;

Query Match 99.4%; Score 6188.5; DB 6; Length 1189;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

```
QY 1 MDLPRGLVVAWALSILWPGFTDTFNMDTRKPRVIFGSRITAFPGYTVQOHDISGNKWLTVGA 60
Db |||
QY 61 PLETNGYQKTGDVYKCPVHGNCTKLNILGRVTLNSVSRKDNMRGLSLATPNKDNSFLA 120
Db |||
QY 61 PLETNGYQKTGDVYKCPVHGNCTKLNILGRVTLNSVSRKDNMRGLSLATPNKDNSFLA 120
Db |||
```

QY 121 CSPLWSHECGSSYYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
Db 121 CSPLWSHECGSSYYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
QY 181 VQHFLINILUKKYYIGBQIQGVGVQGVGDVVEHFLNDYRSVKDVVAAASHIEQRGGTET 240
Db 181 VQHFLINILUKKYYIGBQIQGVGVQGVGDVVEHFLNDYRSVKDVVAAASHIEQRGGTET 240
QY 241 RTAFGLEPARSFAFGKGRGKAKKVMIVITDGSHDSPLEKVIQOSERDNTVRAYAVL 300
Db 241 RTAFGLEPARSFAFGKGRGKAKKVMIVITDGSHDSPLEKVIQOSERDNTVRAYAVL 300
QY 301 GYNNRGINPEIFLNEIKYIASDPDDKHFFNVTDEAALKDVIDALGDRIFSLEGTKNKT 360
Db 301 GYNNRGINPEIFLNEIKYIASDPDDKHFFNVTDEAALKDVIDALGDRIFSLEGTKNKT 360
QY 361 SFGLEMSQTGFSSHVVEDGVLLGAVGYDNGAVLKETSAGKVIPLRESYLKEFPEELKN 420
Db 361 SFGLEMSQTGFSSHVVEDGVLLGAVGYDNGAVLKETSAGKVIPLRESYLKEFPEELKN 420
QY 421 HGAYLGYTTSVVSRRQGRVYVAGAPRFNHTGKVLFTMHNNRSLTIHOAMRGQQIGSYF 480
Db 421 HGAYLGYTTSVVSRRQGRVYVAGAPRFNHTGKVLFTMHNNRSLTIHOAMRGQQIGSYF 480
QY 481 GSEITSVDIDGVTDLVLLGAPMYFNEGRGKVVYVELRQNRFYNGTLKDSHSYQNA 540
Db 481 GSEITSVDIDGVTDLVLLGAPMYFNEGRGKVVYVELRQNRFYNGTLKDSHSYQNA 540
QY 541 RFGSSIASVRDLNODSYNDVVGAPLEDNHAGAIYIFHGFRGSILKTPQORTASELATG 600
Db 541 RFGSSIASVRDLNODSYNDVVGAPLEDNHAGAIYIFHGFRGSILKTPQORTASELATG 600
QY 601 LOYFGCSIHGQDLNEDGLIDLAVALGNAVLMSRPVVQINASHLHFEPSKINIFHRDCK 660
Db 601 LOYFGCSIHGQDLNEDGLIDLAVALGNAVLMSRPVVQINASHLHFEPSKINIFHRDCK 660
QY 661 RSGRDATCAAFLCFTPIFLAPHFQTTVIGIRYNATMDERRYTTPRAHLDGEGDRFTNRVAV 720
Db 661 RSGRDATCAAFLCFTPIFLAPHFQTTVIGIRYNATMDERRYTTPRAHLDGEGDRFTNRVAV 720
QY 721 LLSGQELCERINFHVLDIADYVVKVTFVSVEYSLDDPHGPMDDGWPPTLVSVVPFWNG 780
Db 721 LLSGQELCERINFHVLDIADYVVKVTFVSVEYSLDDPHGPMDDGWPPTLVSVVPFWNG 780
QY 781 CNEDEHCVDPDLVDARSDLPTAMEYCORVLRPAQDCSAYTILSDFTTVPILBSTRQVAV 840
Db 781 CNEDEHCVDPDLVDARSDLPTAMEYCORVLRPAQDCSAYTILSDFTTVPILBSTRQVAV 840
QY 841 EATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSECVNEERLQKQCNVSYPPF 900
Db 841 EATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSECVNEERLQKQCNVSYPPF 900
QY 901 RAKAKVAFDLDEFSKSIPLHLELELAGSSNERDSTKEDNVAPLRHLKYEADVLFT 960
Db 901 RAKAKVAFDLDEFSKSIPLHLELELAGSSNERDSTKEDNVAPLRHLKYEADVLFT 960
QY 961 RSSLSHVBVKLNSLDERVDGIPPPSCIFRQNLGLPPIHGMWKITIPATRSNRL 1020
Db 961 RSSLSHVBVKLNSLDERVDGIPPPSCIFRQNLGLPPIHGMWKITIPATRSNRL 1020
QY 1021 KLRFELTDE-ANTSCNIWGNSTYRPTPVEEDLRAPOLNHSNDVVSINCNRLVPQOE 1079
Db 1021 KLRFELTDEVANTSCNIWGNSTYRPTPVEEDLRAPOLNHSNDVVSINCNRLVPQOE 1080
QY 1080 INFHLLGNLWLSLKLKLYKMKIMVNAALQCFHSPTIFREDEPSROLEFISKQEDWQ 1139
Db 1081 INFHLLGNLWLSLKLKLYKMKIMVNAALQCFHSPTIFREDEPSRQIVFISKQEDWQ 1140
QY 1140 VPIWIVGSLGGLLALLVLLALRLKLGFFRSARRRREFGLDTPKVL 1188
Db 1141 VPIWIVGSLGGLLALLVLLALRLKLGFFRSARRRREFGLDTPKVL 1189

RESULT 11
ADE63570
ID ADE63570 standard; protein; 1189 AA.
XX
AC ADE63570;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human Protein Q9UKX5, SEQ ID NO 9514.
XX
KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; Chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
OS Homo sapiens.
XX
FN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
CC The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1189 AA;

Query Match 99.4%; Score 6188.5; DB 7; Length 1189;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDLPRGLVAVALSMPGFTDTFNNMDTRKPRVPGSRTAFFGYTVOQHDISGNKWLIVGA 60

Db 1 MDLPRGLVVAWALS... 60
Qy 61 PLENGYQKTDVYKCPVHGNCTKMLGRVTLNSVSRKDNMRLGLSLATPNKDSFLA 120
Db 61 PLENGYQKTDVYKCPVHGNCTKMLGRVTLNSVSRKDNMRLGLSLATPNKDSFLA 120
Qy 121 CSPLWSHECGSSVYTCMCSRVSNSRFSKTVAPALORCOTYMDIVIVLQGSNSIYPWVE 180
Db 121 CSPLWSHECGSSVYTCMCSRVSNSRFSKTVAPALORCOTYMDIVIVLQGSNSIYPWVE 180
Qy 181 VQHEFLINILKFFYIGPGQIQGVVQYQGVDDVHVEFHLNDYRSVKDVVEAASHIEQGGTET 240
Db 181 VQHEFLINILKFFYIGPGQIQGVVQYQGVDDVHVEFHLNDYRSVKDVVEAASHIEQGGTET 240
Qy 241 RTAFGIETSEAFQKGRKGAKKVMIVITDGHSDPDLKVIQQSERDNTVRYAVAVL 300
Db 241 RTAFGIETSEAFQKGRKGAKKVMIVITDGHSDPDLKVIQQSERDNTVRYAVAVL 300
Qy 301 GYNNRGINPETFLNEIKYIASDDPKHFNVTDEALKQIVDALGRIFSLGCTNKNET 360
Db 301 GYNNRGINPETFLNEIKYIASDDPKHFNVTDEALKQIVDALGRIFSLGCTNKNET 360
Qy 361 SFGLEMSQTFSSHWEDGVLGAVGAYDMNGAVLKETSAKVIPLRESYLKEPPEELKN 420
Db 361 SFGLEMSQTFSSHWEDGVLGAVGAYDMNGAVLKETSAKVIPLRESYLKEPPEELKN 420
Qy 421 HGAYLGYTVTSVSSRQGRVVVAGAPRNHTGKVLFTMHNNSRLTIHQAMRGQOIGSYF 480
Db 421 HGAYLGYTVTSVSSRQGRVVVAGAPRNHTGKVLFTMHNNSRLTIHQAMRGQOIGSYF 480
Qy 481 GSEITSDVIDGQVTDVLLVCAPMYFNEGRGKVVYVELQNRFPVYNGTLKDSHSYQNA 540
Db 481 GSEITSDVIDGQVTDVLLVCAPMYFNEGRGKVVYVELQNRFPVYNGTLKDSHSYQNA 540
Qy 541 RFGSSIASVRDLNDSYNDVVVGAFLPDNDHAGAYIYFHGFRGSILTKPKQITASELATG 600
Db 541 RFGSSIASVRDLNDSYNDVVVGAFLPDNDHAGAYIYFHGFRGSILTKPKQITASELATG 600
Qy 601 LOYFGCSIHGOLDLNEDEGLIDAVGALGNVILMSRPVQVNASLHPEPSKINIFHRDCK 660
Db 601 LOYFGCSIHGOLDLNEDEGLIDAVGALGNVILMSRPVQVNASLHPEPSKINIFHRDCK 660
Qy 661 RSGRDATCLAAFLCFTPIFLAPHQTTTVGRYNATMDERYTPRAHLDGGRDFTNRAV 720
Db 661 RSGRDATCLAAFLCFTPIFLAPHQTTTVGRYNATMDERYTPRAHLDGGRDFTNRAV 720
Qy 721 LLSGQELCERINFHVLDTADYVKPVTFSVEYSLEDPDHGMPLDDGNPTTLRVSVPFWNG 780
Db 721 LLSGQELCERINFHVLDTADYVKPVTFSVEYSLEDPDHGMPLDDGNPTTLRVSVPFWNG 780
Qy 781 CNEDEHCVDPDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTLSPDTTVFIESTRQRAV 840
Db 781 CNEDEHCVDPDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTLSPDTTVFIESTRQRAV 840
Qy 841 BATLENRGENAYSTVNLISQSANIQFASLIQKEDSDGSEICVNERRLQKQVNCVSPFF 900
Db 841 BATLENRGENAYSTVNLISQSANIQFASLIQKEDSDGSEICVNERRLQKQVNCVSPFF 900
Qy 901 RAKAKVAPRLDSEFSKSIFFLHLELELAAGSDSNERSDSTKEDNVAPLRFHLKYEADVLT 960
Db 901 RAKAKVAPRLDSEFSKSIFFLHLELELAAGSDSNERSDSTKEDNVAPLRFHLKYEADVLT 960
Qy 961 RSSLSHVEVKLNSLERYDGIQPPFSCIFRIQNLGLPFIHGMKMTIPIATRSNRL 1020
Db 961 RSSLSHVEVKLNSLERYDGIQPPFSCIFRIQNLGLPFIHGMKMTIPIATRSNRL 1020
Qy 1021 KLRFELTDE-ANTSNIWGNSTERYPTVEEDLRAPQINHSNSVWVINCINRLVPNOE 1079
Db 1021 KLRFELTDEVANTSNIWGNSTERYPTVEEDLRAPQINHSNSVWVINCINRLVPNOE 1080
Qy 1080 INFHLLGNLWLSLKALKYKSMKIMVNAALQORPHSPFIIFREDESPQIETFEISKQEDWQ 1139

Db 1081 INFHLLGNLWLSLKALKYKSMKIMVNAALQORPHSPFIIFREDESPQIETFEISKQEDWQ 1140
Qy 1140 VPIWIIIVGSTLGGILLALLVLAIRKLCGFRSARRRREPGLDPTPKVLE 1188
Db 1141 VPIWIIIVGSTLGGILLALLVLAIRKLCGFRSARRRREPGLDPTPKVLE 1189
RESULT 12
ADE86584
ID ADE86584 standard; protein; 1189 AA.
XX ADE86584;
XX AC
XX 29-JAN-2004 (first entry)
XX
DE Novel human secreted protein #7.
XX human; secreted protein; cancer; liver disorder; hepatitis;
XX neural disorder; Alzheimer's disease.
XX Homo sapiens.
XX OS
XX US2003129685-A1.
XX
PD 10-JUL-2003.
XX
PF 18-APR-2001; 2001US-00836353.
XX
PR 28-OCT-1998; 98US-0105971P.
PR 27-OCT-1999; 99WO-US025031.
PR 19-APR-2000; 2000US-0198407P.
XX
XX (NIJJ/) NI J.
PA (YOUN/) YOUNG P E.
PA (KENN/) KENNY J J.
PA (OLSE/) OLSEN H S.
PA (MOOR/) MOORE P A.
PA (WEIY/) WEI Y.
PA (GREE/) GREENE J M.
PA (RUBE/) RUBEN S M.
XX
PI Ni J, Young PE, Kenny JU, Olsen HS, Moore PA, Wei Y, Greene JM;
Ruben SM;
XX
XX WPI; 2004-020335/02.
DR N-PSDB; ADE86566.
XX
PT New nucleic acid molecule, useful for preparing a medicament for
preventing, treating or ameliorating a medical condition e.g. cancer,
liver disorders or neural disorders.
XX
PS Claim 11; SEQ ID NO 35; 380pp; English.
XX
CC The invention relates to an isolated nucleic acid sequence, or its
allelic variant, a fragment of the cDNA sequence, or its fragment,
domain, epitope or species homologue. The nucleic acid is useful for
preparing a medicament for preventing, treating or ameliorating a medical
condition e.g., cancer, liver disorders such as hepatitis or neural
disorders such as Alzheimer's disease. The present sequence represents
the amino acid sequence of a novel human secreted protein.
XX
SQ Sequence 1189 AA;
Query Match 99.4%; Score 6188.5; DB 8; Length 1189;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
Qy 1 MDLPRGLVVAWALS... 60
Db 1 MDLPRGLVVAWALS... 60
Qy 61 PLENGYQKTDVYKCPVHGNCTKMLGRVTLNSVSRKDNMRLGLSLATPNKDSFLA 120

CC		may be used to diagnose, treat or prevent metabolic disorders such as
CC		diabetes or obesity, infections, cachexia, cancer, neurodegenerative
CC		disorders such as Alzheimer's disease or Parkinson's disease, immune
CC		disorders, haematopoietic disorders and various dyslipidaemias
XX		
SQ	Sequence 1120 AA;	
	Query Match	92.9%; Score 5780; DB 6; Length 1120;
	Best Local Similarity	93.8%; Pred. No. 0;
	Matches 1115; Conservative	1; Mismatches 3; Indels 70; Gaps 2;
QY	1 MDLPRGLVAVAWLSLWPGFTDTFNMDTRKPRVIPGSRTPAFGYTVQQHDISGNKWLVVGA	60
Dd	1 MDLPRGLVAVAWLSLWPGFTDTFNMDTRKPRVIPGSRTPAFGYTVQQHDISGNKWLVVGA	60
QY	61 PLETNGYOKTGDVVKCPVIHGNCTKLNLRVTLSNVSEKDNMRGLSLATNPKDNGFLA	120
Dd	61 PLETNGYOKTGDVVKCPVIHGNCTKLNLRVTLSNVSEKDNMRGLSLATNPKDNGFLA	120
QY	121 CSP.LWSHECGSSYYTTCGCSRVSNNFRFSKTVPALORCQTYMDIVILDGNSNIYPWVE	180
Dd	90 -----CQTYMDIVILDGNSNIYPWVE	111
QY	181 VQHFLINILKKFYIGPQIQVGVOYGEDVVBHPLNDYRSVKDVAAASHIEORGGETET	240
Dd	112 VQHFLINILKKFYIGPQIQVGVOYGEDVVBHPLNDYRSVKDVAAASHIEORGGETET	171
QY	241 RTAFGIEFARSEAFQKGRKGAKVMITVDGESHDSPDLKEKVQQSERDNVTRYAVAVL	300
Dd	172 RTAFGIEFARSEAFQKGRKGAKVMITVDGESHDSPDLKEKVQQSERDNVTRYAVAVL	231
QY	301 GYNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAALKOIVDALGDRIFSLEGTNKNET	360
Dd	232 GYNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAALKOIVDALGDRIFSLEGTNKNET	291
QY	361 SFGLEMSQTGFSSHVVDEGVLLGAVGDWNGAVLKETSAGKVLPLRESYLKEPPEELKN	420
Dd	292 SFGLEMSQTGFSSHVVDEGVLLGAVGDWNGAVLKETSAGKVLPLRESYLKEPPEELKN	351
QY	421 HGAVLGYTVTSVSSROGRVAVGAPRNFHTGKVLFTFMHNRSLTTHQAMRGQOIGSYF	480
Dd	352 HGAVLGYTVTSVSSROGRVAVGAPRNFHTGKVLFTFMHNRSLTTHQAMRGQOIGSYF	411
QY	481 GSEITSVIDIDGGVTDVLLVGAPMYFNEGREGKVVYVELQRNRPVYNGTLKDHSYQNA	540
Dd	412 GSEITSVIDIDGGVTDVLLVGAPMYFNEGREGKVVYVELQRNRPVYNGTLKDHSYQNA	471
QY	541 RFGSSIASVRDLNODSNVNDVVGAPLENHAGAIIYFHGPRGSILKTKPORITASLATG	600
Dd	472 RFGSSIASVRDLNODSNVNDVVGAPLEDNHAGAIIYFHGPRGSILKTKPORITASLATG	531
QY	601 LQYFGCSIHGOLDLNEGLIDLAVCALGNAILWSRPVVQLNASIHFEPSKINIFHRDCK	660
Dd	532 LQYFGCSIHGOLDLNEGLIDLAVCALGNAILWSRPVVQLNASIHFEPSKINIFHRDCK	591
QY	661 RSGRDATCLAAFLCFTPIFLAPHFQTTTVGIRYNATMDERRYTPRAHLDEGGDRETNRVA	720
Dd	592 RSGRDATCLAAFLCFTPIFLAPHFQTTTVGIRYNATMDERRYTPRAHLDEGGDRETNRVA	651
QY	721 LLSSGOELCERINFHVLDIADYVKPVTSVSYSLEDDPHGPM.LDGGWP.TTLRVSVPPFWNG	780
Dd	652 LLSSGOELCERINFHVLDIADYVKPVTSVSYSLEDDPHGPM.LDGGWP.TTLRVSVPPFWNG	711
QY	781 CNEDEHCVPD.VLIDARS.DLP.TAMEY.CQRLKP.AQDC.SAY.TLS.FDPT.TTFVII.ESTEQ.RVAV	840
Dd	712 CNEDEHCVPD.VLIDARS.DLP.TAMEY.CQRLRKPA.QDC.SAY.TLS.FDPT.TTFVII.ESTEQ.RVAV	771
QY	841 EATTENRGENAYSTVLNISQSANLQFASLIQKEDSDSGISECVNEBRRLQKQVCNVSYPFF	900
Dd	772 EATTENRGENAYSTVLNISQSANLQFASLIQKEDSDSGISECVNEBRRLQKQVCNVSYPFF	831
QY	901 RAKAKVAPRLDSFEFSKSIFLHHLETLAAGSDSNERSDSTKEDNVAPLFHLKYEADV.LFT	960

XX PI Pan Y, Lora JM;
 XX DR WPI; 2001-041142/05.
 XX N-PSDB; AAC91904, AAC91905.
 XX PT Nucleic acid encoding alpha-integrin subunits, useful for treatment and
 XX diagnosis of fibrosis, e.g. of the liver.
 XX PS Claim 8; Fig 5; 164pp; English.
 XX CC The present sequence is murine integrin alpha subunit, A259. A259 is
 CC homologous with the alpha1 and alpha10 integrin subunits and is
 CC overexpressed in fibrosis. A259 is implicated in regulation of
 CC proliferation, differentiation and/or function of many different cell
 CC types. Inhibitors of A259 activity are useful for the treatment of liver
 CC disease, particularly fibrosis, and also fibrosis in other organs
 CC (specifically lung and kidney). In addition, A259 can be used for
 CC treatment and prevention of cancer, osteoporosis, acute myeloid
 CC leukaemia, HIV infection, and rheumatoid arthritis
 XX CC Sequence 1188 AA;
 XX SQ
 Query Match 91.5%; Score 5693; DB 4; Length 1188;
 Best Local Similarity 90.2%; Pred. No. 0;
 Matches 1071; Conservative 57; Mismatches 60; Indels 0; Gaps 0;
 QY 1 MDLPRLVAVALSILWPGTDFNMTRPRVIPSRTAFPGYTVQOHDISGNKMLVGA 60
 Db 1 MDLPRLVAVALSILWPGTDFNMTRPRVIPSRTAFPGYTVQOHDISGNKMLVGA 60
 QY 61 PLETHGYQKTDGVYKCPVTHGCTKLNLRVTLNVSRKDNRLGLSLATNPKNSFLA 120
 Db 61 PLETHGYQKTDGVYKCPVTHGCTKLNLRVTLNVSRKDNRLGLSLATNPKNSFLA 120
 QY 121 CSPLMSHECGSYTTGMCSSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
 Db 121 CSPLMSHECGSYTTGMCSSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
 QY 181 VOHFLINILKPYIIGPGQIQGVQYGEDVHFEHLNDYRSKDVVEAASHIEQGGTET 240
 Db 181 VOHFLINILKPYIIGPGQIQGVQYGEDVHFEHLNDYRSKDVVEAASHIEQGGTET 240
 QY 241 RTAFGIEFARSAFQKGRKGAKKVMIVITDGESHDSPLEKVIQOOSRDNTRVAVL 300
 Db 241 RTAFGIEFARSAFQKGRKGAKKVMIVITDGESHDSPLEKVIQOOSRDNTRVAVL 300
 QY 301 GYNRRGINPETFLNEIKYIASDPDDKHFFNVTDSEAALKDVIDALGDRIFSLEGTNKNET 360
 Db 301 GYNRRGINPETFLNEIKYIASDPDDKHFFNVTDSEAALKDVIDALGDRIFSLEGTNKNET 360
 QY 361 SFGLEMSQTGFSSHVVDEGVLGAYDNGAVLKETSAGKVIPLRESYLKEPPEELKN 420
 Db 361 SFGLEMSQTGFSSHVVDEGVLGAYDNGAVLKETSAGKVIPLRESYLKEPPEELKN 420
 QY 421 HGAYLGTYVTSSVRQGRVYVAGAPRNFHTGKVLFTMNNRSLTIHQAMRGOQIGSYF 480
 Db 421 HAAYLGTYVTSSVRQGRVYVAGAPRNFHTGKVLFTMNNRSLTIHQALRGEQIGSYF 480
 QY 481 GSEITISVDIDGCVTDVLLVGAPMYPNEGRBKGVYVELRQNRFFVNGTLDKDSHYQNA 540
 Db 481 GSEITISVDVNDVRVTDVLLVGAPMYPFSEGREGRKVYVNLQRNFVYNGTLDKDSHYQNA 540
 QY 541 RFGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIFHFGFSGIILKTPKORITASELATG 600
 Db 541 RFGSCIASVQDLNQDSYNDVVVGAPLEDSHAGAIYIFHGFQNLILKPKMQRITASELAPG 600
 QY 601 LQYFCCSHGQDLNEDGLIDLAVGALGNVILWSRPVVQINASLHFFPFSKINIEHRDCK 660
 Db 601 LQHFCCSHGQDLNEDGLVDLAVGALGNVILWARPVVQINASLHFFPFSKINIEHCK 660
 QY 661 RSGRATCLAAFLCFPIPLAFPHFQTATVIGRYNATMDERRYTPRAHLDGEGDRTNRAV 720
 Db 661 RSGRATCLAAFLCFPIPLAFPHFQTATVIGRYNATMDERRYTPRAHLDGEGDRTNRAV 720

Db 661 RSGRATCLAAFLCFPIPLAFPHFQTATVIGRYNATMDERRYTPRAHLDGEGDRTNRAV 720
 QY 721 LLSGQELCERINFHVLDADYVVKPVTFSVEYSLBEDPHGPMDDGWPTTLRVSVDFWNG 780
 Db 721 LLSGQEHQCRINFHVLDADYVVKPVTFSVEYSLBEDPHGPMDDGWPTTLRVSVDFWNG 780
 QY 781 CNEDEHCVDPDLVLDARSDLPTAMEYCORVLRPAQDCSAYTTSFTTTFVIEESTRORVAV 840
 Db 781 CNEDEHCVDPDLVLDARSDLPTAMEYCORVLRPAQDCSAYTTSFTTTFVIEESTRORVAV 840
 QY 841 EATLENRGENAYSTVLNISQSANLOFASLIQKEDSDSGSIECVNEERRLOKQVCNVSYPFF 900
 Db 841 EATLENRGENAYSAVLNISQSSENLOFASLIQKDDSDNSIECVNEERRLHKVCNVSYPFF 900
 QY 901 RAKAKVAFRLDSEFSKSIPLHLETELAAGSOSNERDSTKEDNVAPLRPHLYEADVLT 960
 Db 901 RAKAKVAFRLDSEFSKSVFLHLQHLGAGSOSHEQSDTADNTALLRPHLYEADVLT 960
 QY 961 RSSLSHVEVKNLSLERYDGIKPPFCIFRQNLGLRPIHGMKMITIPIATRSNRL 1020
 Db 961 RSSLSHVEVKNLSLERYDGIKPPFCIFRQNLGLRPIHGMKMITIPIATRSNRL 1020
 QY 1021 KLRFELTDEANTSCNIGWNSTEYRPTVVEEDLRAPQLMHSNSDVVSNCLRLVNPQBI 1080
 Db 1021 MLRDFPTQNTSCNIGWNSTEYRSTPTEEDLSHAPQNHNSDVVSIICNLRLAPSQT 1080
 QY 1081 NFHLGNLWLSLKALKYKMKIMVNAALQRFHSPFFIFREDDPSRQIEFEISKQEDWQV 1140
 Db 1081 SFYLVGNLWLSLKALKYKSLKITVNAALQRFHSPFFIFREDDPSRQIEFEISKQEDWQV 1140
 QY 1141 PIWIIVGSTGLGLLLALLVIALRLKLGFFRSARRRREGLDPTPKVLE 1188
 Db 1141 PIWIIVGSTGLGLLLALLVIALRLKLGFFRSARRRREGLDPTPKVLE 1188
 RESULT 15
 AAU10552
 ID AAU10552 standard; protein; 1188 AA.
 XX AC AAU10552;
 XX DT 14-FEB-2002 (first entry)
 XX Murine A259 polypeptide.
 DE Mouse; A259; integrin alpha subunit; integrin alpha 10; secreted protein;
 KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;
 KW immune related disease; apoptotic disorder; neuronal tissue disease;
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
 KW antithratic; antinaemic; antiallergic; antidiabetic; dermatological;
 KW antidiabetic; anticonvulsant; antiparkinsonian.
 OS Mus musculus.
 XX Key Location/Qualifiers
 FH Domain 1..1141
 FT /note= "Extracellular domain"
 FT Peptide 1..22
 FT /note= "Signal peptide"
 FT Protein 23..1188
 FT /note= "Mature murine A259"
 FT Domain 39..74
 FT /note= "Integrin alpha repeat domain"
 FT Domain 115..157
 FT /note= "Integrin alpha repeat domain"
 FT Domain 164..345
 FT /note= "I domain or Von Willebrand Factor type A domain"
 FT Domain 367..392
 FT /note= "Integrin alpha repeat domain"
 FT Domain 421..455
 FT /note= "Integrin alpha repeat domain"

FT	Domain	478...516	/note= "Integrin alpha repeat domain"
FT	Domain	540...575	/note= "Integrin alpha repeat domain"
FT	Domain	602...640	/note= "Integrin alpha repeat domain"
FT	Domain	1142...1184	/note= "Transmembrane domain"
FT	Domain	1165...1188	/note= "Cytoplasmic domain"
XX	WC200181414-A2.		
XX	01-NOV-2001.		
XX	27-APR-2001; 2001WO-US013516.		
XX	27-APR-2000; 2000US-00561263.		
XX	(MILL-) MILLENNIUM PHARM INC.		
XX	Pan Y, Lora J;		
XX	WPI; 2002-041397/05.		
XX	N-PSDB; AAS16874.		
XX	New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases.		
XX	Claim 9; Fig 5; 168pp; English.		
XX	The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and haematopoietic disorders (such as acute myeloid leukaemia, haemophilia, anaemia and thalassaemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the neuronal tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the murine A259 polypeptide		
XX	Sequence 1188 AA;		
Qy	Query Match	91.5%;	Score 5693; DB 5; Length 1188;
Db	Best Local Similarity	90.2%;	Pred. No. 0;
Qy	Matches 1071; Conservative	57; Mismatches	60; Indels 0; Gaps 0
Qy	1 MDPLRGLVVAWALSLWPGFTTDFNMTRKPRVPGSRTAFFGYTVOQHDISGNKWLVVGA	60	
Db	1 MDFPRGLVVAWTLSLWPGFTTDFNMTRNPRVIAFGPSAAFFGYTVOQHDISGNKWLVVGA	60	
Qy	61 PLENTNGYQKTGDVVVKCPVIHGNCTKLNLRVTLNSVSRKNNMRGLSLATNPKDNSFLA	120	
Db	61 PWEINGHQKTGDVVKCPVTQGNCTKLNLRVTLNSVSRKNNMRGLSLATNPKDNSFLA	120	
Qy	121 CSPLWSHECGSSYYTTCMCSRNSNFRFSKTVAPALQRCQTYMDIVIVLDGSNSIYPWVE	180	
Db	121 CSPLWSHECGSSYYTTCMCSRNSNFRFSKTVAPALQRCQTYMDIVIVLDGSNSIYPWVE	180	
Qy	181 VQHELINILKKFYIGPGQIQGVQVQYGEDVVEHFHLNDYRSVKQVWEAASHTEORGGTET	240	
Db	181 VQHELINILKKFYIGPGQIQGVQVQYGEDVVEHFHLNDYRSVKQVWEAASHTEORGGTET	240	

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:44:02 ; Search time 34.8436 Seconds
(without alignments)
1775.344 Million cell updates/sec

Title: US-09-980-403-2

Perfect score: 6224

Sequence: 1 MDLPRGLVVAWALSLWPGFT.....PSRARRRPGLDTPKVLV 1188

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6188.5	99.4	1189	ITAH_HUMAN	Q9UKX5 homo sapien
2	2403.5	38.6	1167	ITAG_HUMAN	O75578 homo sapien
3	2159	34.7	1151	ITAI_HUMAN	P56199 homo sapien
4	2149.5	34.5	1180	ITAI_RAT	P18614 rattus norv
5	1910.5	30.7	1178	ITR2_MOUSE	Q62469 mus musculu
6	1901.5	30.6	1170	ITR2_BOVIN	P53710 bos taurus
7	1863	28.9	1181	ITR2_HUMAN	P17301 homo sapien
8	1146.5	18.4	1170	ITAD_HUMAN	P20701 homo sapien
9	1139	18.3	1162	ITAD_HUMAN	Q13349 homo sapien
10	1132.5	18.2	1163	ITAX_HUMAN	P20702 homo sapien
11	1104	17.7	1153	ITAM_MOUSE	P05555 mus musculu
12	1093.5	17.6	1152	ITAM_HUMAN	P12125 homo sapien
13	1073.5	17.2	1163	ITAL_MOUSE	P24063 mus musculu
14	954	15.3	1167	ITAE_MOUSE	Q60677 mus musculu
15	924.5	14.9	1179	ITAE_HUMAN	P38570 homo sapien
16	825	13.3	825	ITAI_CHICK	Q90615 gallus gall
17	799.5	12.8	1035	ITR9_HUMAN	Q13797 homo sapien
18	693.5	11.1	1039	ITR4_MOUSE	Q06651 mus musculu
19	673.5	10.8	1032	IT4_KENLA	Q91687 xenopus lae
20	672.5	10.8	1038	IT4_HUMAN	P13612 homo sapien
21	611.5	9.8	1130	ITAG_HUMAN	P23229 homo sapien
22	607.5	9.8	1034	ITAV_CHICK	P26008 gallus gall
23	606	9.7	1049	ITR5_HUMAN	P08648 homo sapien
24	604.5	9.7	1053	ITR5_MOUSE	P11688 mus musculu
25	597.5	9.6	1066	IT3_HUMAN	P26006 homo sapien
26	594	9.5	1053	IT3_MOUSE	Q62470 mus musculu
27	584.5	9.4	1044	ITAV_MOUSE	P43406 mus musculu
28	580.5	9.3	1066	IT3_CRISP	P17852 cricetidae
29	579	9.3	1091	ITR6_MOUSE	Q61739 mus musculu
30	577.5	9.3	1050	ITR5_KENLA	Q06274 xenopus lae
31	568	9.1	1072	ITR6_CHICK	P26007 gallus gall
32	566.5	9.1	1146	ITAI_DROME	Q24247 drosophila
33	563	9.0	1044	ITR8_CHICK	P26009 gallus gall

34 549.5 8.8 1000 1 ITA5_DROME
35 546.5 8.8 1226 1 PAT2_CAEEL
36 542 8.7 1115 1 ITA3_DROME
37 537 8.6 1048 1 ITAV_HUMAN
38 525.5 8.4 1181 1 ITA7_HUMAN
39 520.5 8.4 1179 1 ITA7_MOUSE
40 515.5 8.3 1139 1 INAI_CAEEL
41 509.5 8.2 1025 1 ITA8_HUMAN
42 502.5 8.1 1106 1 ITA7_RAT
43 487 7.8 1033 1 ITAB_MOUSE
44 485.5 7.8 1015 1 ITA4_DROME
45 472.5 7.6 1039 1 ITAB_HUMAN

ALIGNMENTS

RESULT 1
ITAH_HUMAN
ID ITAH_HUMAN STANDARD; PRT; 1189 AA.
AC Q9UKX5; Q9UKQ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-11 precursor.
GN ITGAl1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal heart, and Osteoblast;
RX MEDLINE=99417678; PubMed=10486209;
RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
RA Wang S.-X., Morris C.M., Krissansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel
RT human integrin alpha11 subunit (ITGAl1).";
RN Genomics 60:179-187(1999).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal muscle, and Uterus;
RX MEDLINE=99395147; PubMed=10464311;
RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;
RT "cDNA Cloning and Chromosomal Localization of Human alpha(11)
RT Integrin. A collagen-binding, i domain-containing, beta(1)-associated
RT integrin alpha-chain present in muscle tissues.";
RN J. Biol. Chem. 274:25735-25742(1999).
[3]
RP SEQUENCE OF 954-1188 FROM N.A.
RC TISSUE=Fibroblast;
RA Andreu N., Bativill X., Escarceller M., Sumoy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC !- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
CC !- ASSOCIATES WITH BETA-1.
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
CC !- HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
CC !- LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
CC !- FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS, IN
CC !- SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
CC !- PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
CC !- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
CC !- FETAL MUSCLE CELLS (IN VITRO).
CC !- DOMAIN: THE INTEGRIN I-DOMAIN (INSET) IS A VWFA DOMAIN. INTEGRINS
CC !- WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC !- SIMILARITY: Belongs to the integrin alpha chain family.
CC !- SIMILARITY: Contains 1 VWFA domain.
CC !- SIMILARITY: Contains 7 FG-GAP repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

QY 901 RAKAKVAFRLDSEFSGKSIPLHLELELAAGSDSNEDSTKEDNVAFLRPHLYEADVLFT 960
 DB 901 RAKAKVAFRLDSEFSGKSIPLHLELELAAGSDSNEDSTKEDNVAFLRPHLYEADVLFT 960
 QY 961 RSSLSHVEKVNLSLRYDGTGPPPSCFIRIONTGLFFIHGMKMTIPIATRSNRL 1020
 DB 961 RSSLSHVEKVNLSLRYDGTGPPPSCFIRIONTGLFFIHGMKMTIPIATRSNRL 1020
 QY 1021 KLRFDTDE-ANTSNCNWSNTEYRPTPVEEDLRAPOLNHSNVDVSNICNIRLVNQE 1079
 DB 1021 KLRFDTDE-ANTSNCNWSNTEYRPTPVEEDLRAPOLNHSNVDVSNICNIRLVNQE 1079
 QY 1080 INFLLGNLWLSLKALKYKMKIMVNAALQRFSPFFREDPSRQIEFISKQEDWQ 1139
 DB 1081 INFLLGNLWLSLKALKYKMKIMVNAALQRFSPFFREDPSRQIEFISKQEDWQ 1140
 QY 1140 VPIWIVGSTLGGLLALLLVIALRLKLGFRSARRRREGLDPTPKVLE 1188
 DB 1141 VPIWIVGSTLGGLLALLLVIALRLKLGFRSARRRREGLDPTPKVLE 1189

RESULT 2

ITAG.HUMAN
 ID ITAG.HUMAN STANDARD; PRT; 1167 AA.
 AC 075578; Q9UHZ8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Integrin alpha-10 precursor.
 GN ITGA10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Articular chondrocytes;
 RX MEDLINE=98352078; PubMed=9685391;
 RA Camper L., Hellman U., Lundgren-Akerlund E.;
 RT "Isolation, cloning, and sequence analysis of the integrin subunit
 RT alpha10, a beta1-associated collagen binding integrin expressed on
 RT chondrocytes.";
 RL J. Biol. Chem. 273:20383-20389 (1998).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Endothelial cells, and Heart;
 RX MEDLINE=20169197; PubMed=10702680;
 RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,
 RA Wang S.-X., Langley R., Krissansen G.W.;
 RT "The integrin alpha10 subunit: expression pattern, partial gene
 RT structure, and chromosomal localization.";
 RL Cytogenet. Cell Genet. 87:238-244 (1999).
 CC -!- FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10
 CC ASSOCIATES WITH BETA-1.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Widely expressed with highest expression in
 CC muscle and heart. Found in articular cartilage.
 CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -!- SIMILARITY: Contains 1 VWFA domain.
 CC -!- SIMILARITY: Contains 7 FG-GAP repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AF074015; AAC31952.1; -;
 DR EMBL; AF112345; AAF21944.1; -;
 DR EMBL; AF172723; AAF61638.1; -;
 DR HSSP; P17301; IAOX.
 DR Genew; HGNC:6135; ITGA10.
 DR MIM; 604042; -;
 DR GO; GO:0008305; C:integrin complex; TAS.
 DR GO; GO:0004895; F:cell adhesion receptor activity; TAS.
 DR GO; GO:0005518; F:collagen binding; TAS.
 DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWFA.
 DR Pfam; PF01839; FG-GAP; 3.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 4.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
 DR PROSITE; PS00234; VWFA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Repeat; Calcium; Magnesium.
 FT SIGNAL 1 22
 FT CHAIN 23 1167
 FT DOMAIN 23 1122
 FT TRANSMEM 1123 1145
 FT DOMAIN 1146 1167
 FT REPEAT 38 97
 FT REPEAT ? ?
 FT DOMAIN 167 350
 FT REPEAT 365 427
 FT REPEAT 428 482
 FT REPEAT 483 545
 FT REPEAT 546 605
 FT REPEAT 608 660
 FT DOMAIN 1134 1140
 FT CA_BIND 494 502
 FT CA_BIND 558 566
 FT CA_BIND 620 628
 FT DISULFID 76 86
 FT DISULFID 666 675
 FT DISULFID 681 736
 FT DISULFID 789 795
 FT CARBOHYD 98 98
 FT CARBOHYD 234 234
 FT CARBOHYD 336 336
 FT CARBOHYD 364 364
 FT CARBOHYD 733 733
 FT CARBOHYD 763 763
 FT CARBOHYD 839 839
 FT CARBOHYD 921 921
 FT CARBOHYD 1011 1011
 FT CARBOHYD 1018 1018
 FT CARBOHYD 1039 1039
 FT CONFLICT 844 844
 FT CONFLICT 909 909
 FT CONFLICT 926 926
 SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AE0 CRC64;

Query Match 38.6%; Score 2403.5; DB 1; Length 1167;
 Best Local Similarity 43.0%; Pred. No. 6.4e-148;
 Matches 514; Conservative 212; Mismatches 425; Indels 45; Gaps 16;
 QY 1 MDPLRGLVAVAWLSLWFGFTDTFNMMDTRKPRVIFGSRFAFGYTVQOHDISGNKWLVGA 60
 DB 1 MELPFVTHLFLPLVFLTGLCSFENLDEHHPGLPPEAEFGYSLQHVGGQRWMLVGA 60
 QY 61 PLETFNGYQKGVGVYKCVY---IHGNCCTKLMLRVTLSNVRKDNMLGLSLATNPKNDS 117
 DB 61 PWDGSGDRRGDVYRCVPGVGAHNAAPCAKHLGDYQLGNSHHPAVNMHLGMSLLETGDDG 120
 QY 118 FLACSPLSWSECGSGSYTTGTGCSRVNSNFRFSKTVAPALORCQTYMDIVIVLDGNSNYP 177

Db 121 FMACAPLWSRACGSSVSSGICARVDASFPQGSGLAPTAQRCTYMDVVIVLDGNSIYP 180
QY 178 WVEVQHLINILKFKYIGPQIQGVVQYGEDVWHEFHLDNYSRVKDVVAAASHIBQRGG 237
Db 181 WSEVQFRLRLVGLKFLDPEQIQVGLVQGESVHWSLGDPRTEKEVVRRAAKNLSRREG 240
QY 238 TETRTAFGTEFARSEAFQK--GGRKCAKVMIVITDGHSDPDLKVKVQOSRDWTRY 295
Db 241 RTTKTAAIMVACTEGFSQSHGRPEAARLLVVVTDGESHDEELPAALKACEAGVTRY 300
QY 296 AVAVLGYNNRGINPTTFINEIKYIASDPDDKHFFNVTDAAIKOIVDAGDRIFLEGT 355
Db 301 GTAVLGHYLRQRDPSFIREIRTIASDDPDRFFNVTDAAITDIVDAGDRIFGLEGS 360
QY 356 N-KNETSFGLMSQTSFSSHVEDGVLLGAVGAYDNGAVLKETSAGKVIPIRESYKPF 414
Db 361 HAENESSFGLMSQIGFSTHRLKDGILFGVGVAYDNGSVLWEGGHRLLFPPEMALEDEF 420
QY 415 PEELKNGAYLVYTVTSVVSSRQGVVYVAGAPREFNHTGVLLPTMNNBSLTHQAWRQ 474
Db 421 PPALQNHAAAYLVGSVSSMLLRGRRLLFGSAPRFRHGRKVIAFQPKKDGAVRVAQSLQGE 480
QY 475 QIGSYFSGSHTSDIDGDGVDVLLVGAPEYFN-EGRERGKVVYVEL-RQNRVYVNGTLK 532
Db 481 QIGSYFSELCLDTRDGTVDVLLVAAPFLGPQNKETGRVYVVLVGGQSLTLTLOGTLQ 540
QY 533 DSHSYQNAFSGSIASVRDLNQDSYNDVVVGAPELDNHAAYIYIFGFRGSIUKTPQRI 592
Db 541 PEPP-QDARFGFAMGALPDNLQGFADVAVGAPLEDHQALYVHGTQSGVRPFPQRI 599
QY 593 TASELATGQYFGCSITHGOLDNEDGLDLAVCALNAVILWRPVVQVQNASLHFPSKI 652
Db 600 AAASMPHALSYFORSVDGRDLDDGDLVDVAVAGAAILLSRPIVHLTPLEVTPOAI 659
QY 653 NIFHRCKSGSRDATCLAAFLCTPIFLAPHFQTTTVGIRYNATMDERRYPRAHDEGG 712
Db 660 SVVQRCRRRQGEAVCLTAALCFQVTSRTPTGRWDHQFYMYFTASLDWTAGARAAPDGG 719
QY 713 DRPTNRAVLLSSQELCERINFHVLDTADYVVKVPTSVEYSLSDPDH-GRMLDDGHPPTL 771
Db 720 QRLSPRLRLSVGNVTCEQLHFEHVLDTSDYLRVAVLTVPALDNTTKPGFVLNCGSPSI 779
QY 772 RUSVPVWNGNEDEHCVPLDLARSDLPFAMEYCORVLRKPAQDCSAYLSTPDTTFVII 831
Db 780 QKLVPSKCGPNECVTDLVLVQNMDI-----RGRK-----APFVV 817
QY 832 ESTQRVAVATLENRGENAYSTVINISQANLQFASLIQKEDSGSIECNBERLQKQ 891
Db 818 RGRKRVLVSTLENKENAYNTSLIIPSRNLHLASLTPQRESPIKVECAAPS--AHAR 875
QY 892 VCVSVYFPFRKAKVAFRLDSEFSKSIHLHLBIELAAGSDSNERSDKEDNAPLRFHL 951
Db 876 LCSVGHFVFTGAKVTLLEFEFESCSLSLQVFGKLTASSDSLERNGLTQENTAQTSAYI 935
QY 952 KYEADVLFTSSLSHYEVKLSRLERYDYGIPFPSCIFRIQMLGLPFIHGMWKITPI 1011
Db 936 QYEPHLLFSESTLHYEVHPYGTLP--VCPGPEFTTLVQNLGCYVWSGLISALLPA 993
QY 1012 ATRSGNRLKLRLDFTLDEANTSCNMGNSTEYRPTPVE-BDLRRAPOLNHSNDVVSINC 1070
Db 994 VAHGGNVFLSLSQVITN--NASC-IVQNLTEPPGPVHPPEELQHTNRLNGSNTCCQVRC 1050
QY 1071 NI-RLVFNQINPHLLGNLWRLSLKALKYKSMKIMVNAALQORQHPFPFIREEDPSQIE 1129
Db 1051 HLGQLAGTEVSVGLLRLVNEFRFRKFRKSLTVVSTFELGTGEGSVLQTEASRWSESL 1110
QY 1130 FEISKQBDWQVPIIIVGSGTIGGLLALLVLRKLGLGFFRSAR----RRRPGLD 1181
Db 1111 LEVVQTPILLISLWILIGSVLGGILLALLVFLWKLGLGFFAHKKIPKEERREKLE 1166

RESULT 3
ITAL_HUMAN

ITAL_HUMAN STANDARD; PRT; 1151 AA.
P56199;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
ITGAL.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=93155124; PubMed=8428973;
Briesewitz R., Epstein M.R., Marcantonio E.E.;
"Expression of native and truncated forms of the human integrin alpha
1 subunit";
J. Biol. Chem. 268:2989-2996(1993).
CC -!- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-P-P-G-
E-R IN COLLAGEN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD49a entry;
WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd49a.htm".
DR PIR; A45226; A45226.
DR FDB; 1QC5; 17-MAY-00.
DR Genev; HGNC:6134; ITGAL.
DR MIM; 192968; -.
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0004895; F:cell adhesion receptor activity; NAS.
DR GO; GO:0005518; F:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; NAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; VWFA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
Repeat; Calcium; Magnesium; 3D-structure.
FT DOMAIN 1 1113 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1114 1136 POTENTIAL.
FT DOMAIN 1137 1151 CYTOPLASMIC (POTENTIAL).
FT REPEAT 16 75 FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT DOMAIN 147 360 VWFA.
FT REPEAT 349 404 FG-GAP 3.
FT REPEAT 405 457 FG-GAP 4.
FT REPEAT 459 520 FG-GAP 5.
FT REPEAT 540 599 FG-GAP 6.
FT REPEAT 602 654 FG-GAP 7.
FT CA_BIND 470 478 POTENTIAL.
FT CA_BIND 552 560 POTENTIAL.
FT CA_BIND 614 622 POTENTIAL.
FT SITE 1139 1142 GPEKR MOTIF.
FT DISULFID 54 64 BY SIMILARITY.
FT DISULFID 660 669 BY SIMILARITY.
FT DISULFID 675 728 BY SIMILARITY.
FT DISULFID 780 786 BY SIMILARITY.
FT DISULFID 850 858 BY SIMILARITY.
FT DISULFID 1002 1034 BY SIMILARITY.
FT DISULFID 1037 1044 BY SIMILARITY.
FT CARBOHYD 46 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 46

```

FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 812 812 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 855 855 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT CARBOHYD 1085 1085 N-LINKED (GLCNAC. . .) (POTENTIAL) .
SQ SEQUENCE 1151 AA; 127837 MW; 6B3F3C1AABF52808 CRC64;

Query Match 34.7%; Score 2159; DB 1; Length 1151;
Best Local Similarity 38.3%; Pred. No. 4.7e-132;
Matches 457; Conservative 235; Mismatches 417; Indels 84; Gaps 20;

QY 23 FNMTRKPRVIGSRATFGYTVQOHDISGNKWLUVGAPLENGYQKTDGVVKCPVIHGN 82
DB 1 FNVYKNSMTFGSPVDEMGYTVQYENEGKWLIGSLVQPKNRTGVNCKPVGRC 60
QY 83 ---CYKLNIG-RVTLNVSERKDNRLGLSLATNPKNDSFLACSLPWSHECGSSYTTGM 138
DB 61 SLPCVKLDPVNTSIPNVTVEKNMTFGSTLVNTP-NGGFLACGPLYAVRCGHLHYTGI 119
QY 139 CSFVNSNFFSKTVAPALQRCOTYMDIVIVLDSNSIYPWVEVQHLNLIKFFYIGQC 198
DB 120 CSDVSPTFGVNSIAP-VQECSTQLDVIIVLDSNSIYPWDSVTAFLNDLLKRMIDGPK 178
QY 199 IQGVVQVQGEDVVHFHNDYRSVDVVEAASHIQRGGTERTAFGIEFARSEAF-QK 256
DB 179 TQGVIVQVGENVTHFNLNKYSTEVIVAAKIVQGRQMTALGDTARKEAFTAR 238
QY 257 GGRGAKKVMIVITDGESHSDPLEKVIQOQSERDNVTRYAVAVLGYNNRGINPETFLNE 316
DB 239 GARRGVKKVMIVITDGESHSDHRLKKVIQDCEDENIQRFSAIILGYSNRGNLSTKEFVEE 298
QY 317 IKYIASDPDDKHFNVDVDEAALKDIDVALGDRIFSLGKT-NKNETSFGLEMSQTGFSSHV 375
DB 299 IASIASETEPRHFNVDVDELAJVTIVKTIGERIFALEADQSAASFEMEMSQTGFSAHY 358
QY 376 VEDGVLLGAVGAYDNWGAIVLKETSAGKVIPLRESYLVKEPPELKNHGAVLGYVTVSVSS 435
DB 359 SQDWMLGAVGAYDNWGVVWQKASQIILPRNTTFNVSTKKNELASVLGYTVNSATAS 418
QY 436 RQGVYVAGAPRNTGKVLFTMNNRSLTHIQAMRQQIGSYFGSEITSDVIDGDGVT 495
DB 419 SGDVLYIAGQPRYNTGVIIYRMEDG-NIKILQTLSGEQIGSYFGSILTTDIDKDNST 477
QY 496 DVLVLCGAPWYF-NEGRERCKVYVYELRQNFVYNGTLK-----DSHSYQN--- 539
DB 478 DILLVGAPMYGTEKEQGVVYVYALNQTRFYQMSLEPIKQTCSSSRQHNSTCTTENKNE 537
QY 540 ---ARFGSSIASVRLNODSVNDVVVGAPLEDNHAGAIYIFHGFRGSLKTPKORITASE 596
DB 538 PCGARFGTAIAVKDLNLDGFNDIVIGAPLEDHGAIVIHGSKTIRKEVAQRIPSGG 597
QY 597 LATGLQYFGCSHGLDNEGLDIDLAVGALNAVILWSRPVQVNASLHFEPSPKINIFH 656

```

```

DB 598 DGKTLKFFGQSIHGEMDNLGDLTDVTIGLGAALFWSRDVAVVAVKVTWNFPKNVNIQK 657
QY 657 RCKSGSRDATCLAAPLCPTPIPLAPHFTOTTGIRYNATMDERRVTPPAHLDEGDRFT 716
DB 658 KNCHMEGKETVCINATVCFEVLKSGKEDTIYBADLQYRVTLDSLRLQISRSFSFGTQERKV 717
QY 717 NEAVLLSSGOELCERINERHVLDTADYVKPVTSEVYSLEDPOHGPMLDDGWPTTLAVSVP 776
DB 718 QRNITVRKSE--CTHSHFYMLDKHDFQDSVRITLDFNLTDPENGVPVLDSLNSPHEHYIP 775
QY 777 FWNCGNEDEHCVPLDVLDAARSDLPATAMEYQCRVLRKPAQDCSAYTILSFPTTTFIESTRQ 836
DB 776 FAKDCGNKEKCLSDL-----SLHVATTEKOLLIVRSQND 809
QY 837 RVAVATLENROGENAYSTVLNISOSANLOFASL--IQKEDSDGSIECVNEERRLQKQVCN 894
DB 810 KENVSLTVKNTKDSAYNTRITIVHYSPLNVFSGIEAIQKDCSEN-----HNITCK 859
QY 895 VSPYPRFRKAKVAERLDSEFFSKSIFLHLEIELAAGSDSNERDSTKEONVAPLRPHLYKE 954
DB 860 VGYPFLRGEWTFKILFQNTSYLMENVTIYLSATSDBEPPETLSNDNVNISIPVKYE 919
QY 955 ADVLFRSSSLSHYEVKLNSS----LERYDGIGPPFCIFRIQNLGLFPFHGMKMKITIP 1010
DB 920 VGLQYVSSASEYHISIAANEVPEVINGSTEDIGNEINIFYLIRKSGSPMPELKLSISFP 979
QY 1011 IATRSNRLNLPVQINPQLNGLNLSLKALKYKSMKIMVNAALQROPHSFFIFREEDP 1064
DB 980 NMTSGYFVL-YPTGLSSSENANCRPHIFEDPPFSINSKGKMTTSDHLKRGTLDCNTCK 1038
QY 1065 VVSIINCINRLVNPQINPQLNGLNLSLKALKYKSMKIMVNAALQROPHSFFIFREEDP 1124
DB 1039 FATITCNITSSDISOVNVSLL--LMKPTIKSYFSSLNLTIRGEL-RSENASLVLSGNQ 1095
QY 1125 SRQIEFEISKOB-DWQVPIWIIIVGSTLGLLILLALLVLALRKLGFRRARRRR 1176
DB 1096 KRELAIQISKDGLPGRVPLWILLSAFAGLLMLLILALWKIGFFKPKLKK 1148
RESULT 4
ID ITAL RAT STANDARD; PRT; 1180 AA.
AC P18614;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1)
DE (CD49a).
GN ITGA1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90338125; PubMed=2380249;
RA Ignatius M.J., Large T.H., Houde M., Tawil J.W., Barton A.,
RA Esch F., Carbonetto S., Reichardt L.F.;
RA "Molecular cloning of the rat integrin alpha 1-subunit: a receptor
RA for laminin and collagen."
RA J. Cell Biol. 111:709-720(1990).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.
RX MEDLINE=993113197; PubMed=10386626;
RA Nolte M., Pepinsky R.B., Vanyaminov S.Y., Koteliansky V.,
RA Gotwals P.J., Karpusas M.;
RT "Crystal structure of the alphabeta1 integrin I-domain: insights into
RT integrin I-domain function."
RL FEBS Lett. 452:379-385(1999).
CC -!- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC E-R IN COLLAGEN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1

```



```

Db 881 -----QNTCRVGYFFLRAGETVTKIIFQNTGSHLSNALIHLSATSDSEPLESLN 933
QY 942 DNVAFLRFLHKYADVLETRSSLSHYEVKLNSSLERY-----DGIGPPFSCIFRIQNLGL 997
Db 934 DNEVNISIPKVEVGVQFYSSASEHHLSVAANETIPEFINSTEDIGNEINVFYTIKRGH 993
QY 998 FPIHGMMKTIPIATPSGNRLKLRDLTDEANTSCNIWGNSTERYPTFVEE--DLRRA 1055
Db 994 FEMPELQLSIFPNLTADGVPLVPIGWSSD--NVNC-----RPSLEDPPFGINS 1043
QY 1056 POLNHSNSDVV-----SINCRLVQNEINFHLLGNLWLSUKALKYKS 1100
Db 1044 KWTISKSVLKGRTQDCSSSTCGVATITCSLLPSDLSONVSL--LWKPTPIRAHFS 1101
QY 1101 MKIMVNAALQROPHSPFIFREEDPSQIEFESKQB--DMQVPIWIIIVGSTLGLMLALL 1159
Db 1102 LNLTLRGELKSE--NSLTLSSNRKELAIQISKDLGPRVPLWVILLSAFAGLLMLL 1160
QY 1160 VLALRKLGFRRARRR 1176
Db 1161 ILALWKIGFGRPLKK 1177

RESULT 5
ITAZ_MOUSE
ID ITA2_MOUSE STANDARD; PRT; 1178 AA.
AC Q62469; Q62163;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CRA; TISSUE=Lung;
RX MEDLINE=94363406; PubMed=9081899;
RA Edelman J.M.; Chan B.M.; Uniyal S., Onodera H., Wang D.Z.,
RA Damjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;
RT "The mouse VLA-2 homologue supports collagen and laminin adhesion but
RT not virus binding.";
RL Cell Adhes. Commun. 2:131-143 (1994).
RN [2]
RP SEQUENCE OF 450-1178 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=94355691; PubMed=7521231;
RA Wu J.E., Santoro S.A.;
RT "Complex patterns of expression suggest extensive roles for the alpha
RL Dev. Dyn. 199:292-314 (1994).
CC -!- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A COLLAGEN RECEPTOR, BEING
CC RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER CELLS TO
CC COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE EXPRESSION,
CC FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
CC EXTRACELLULAR MATRIX. IT IS ALSO A RECEPTOR FOR LAMININS, COLLAGEN
CC C-PROPEPTIDES AND E-CADHERIN. MICE HOMIZIGOUS FOR A NULL MUTATION
CC IN THE ALPHA-2 DIE VERY EARLY IN EMBRYOGENESIS.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
CC ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; Z29987; CAA82877.1; -
CC DR EMBL; X75427; CAA53178.1; -
CC DR PIR; S44142; S44142.
CC DR HSSP; PI7301; LAOX.
CC DR MGD; MGI:96600; Itga2.
CC DR InterPro; IPR000413; Integrin_alpha.
CC DR InterPro; IPR002035; VWFA.
CC DR Pfam; PF01839; FG-GAP; 3.
CC DR Pfam; PF00357; integrin_A; 1.
CC DR Pfam; PF00092; vwa; 1.
CC DR SMART; SM00191; int_alpha; 4.
CC DR SMART; SM00327; VWFA; 1.
CC DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC DR PROSITE; PS00234; VWFA; 1.
CC KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
CC KW Platelet; Signal; Repeat; Calcium; Magnesium.
CC FT SIGNAL 1 26
CC FT CHAIN 27 1178
CC FT DOMAIN 27 1178
CC FT TRANSMEM 1130 1151
CC FT DOMAIN 1152 1178
CC FT REPEAT 42 100
CC FT REPEAT ? ?
CC FT DOMAIN 185 375
CC FT REPEAT ? ?
CC FT REPEAT 431 483
CC FT REPEAT 485 546
CC FT REPEAT 548 607
CC FT REPEAT 612 664
CC FT CA_BIND 496 504
CC FT CA_BIND 560 568
CC FT CA_BIND 624 632
CC FT SITE 480 482
CC FT SITE 1154 1158
CC FT DISULFID 80 89
CC FT DISULFID 677 734
CC FT DISULFID 786 792
CC FT DISULFID 862 873
CC FT DISULFID 1016 1047
CC FT DISULFID 1052 1057
CC FT CARBOHYD 102 102
CC FT CARBOHYD 109 109
CC FT CARBOHYD 429 429
CC FT CARBOHYD 457 457
CC FT CARBOHYD 472 472
CC FT CARBOHYD 596 596
CC FT CARBOHYD 1054 1054
CC FT CARBOHYD 1071 1071
CC FT CARBOHYD 1078 1078
CC SQ SEQUENCE 1178 AA; 138926 MW; 1F194B9C0240F465 CRC64;

Query Match 30.7%; Score 1910.5; DB 1; Length 1178;
Best Local Similarity 36.2%; Pred. No. 6.6e-116;
Matches 442; Conservative 238; Mismatches 437; Indels 103; Gaps 31;

QY 6 GLVAVAWALSIPWGFDT--TENMDTRKPRVIGSRITAFEGYTVQOHDISGNKMLVVGAPLE 63
Db 8 GALLQLQLMLVQGLINCLAYNVGLPGKIFSGPSEQFGYSVQQLTNPQGKMLLVGSPWS 67
QY 64 TNGYQKTDGVYKCPV--IHGNCTKLN--GRVTLNVSERKDNRLGLSLATNPKDNSFLA 120
Db 68 GFPENRMGDGVYKCPVDLPATCEKLNQNSASISNVTEIKTNMSLGLTLTENPGTGGLT 127
QY 121 CSPLWSHECGSSYTTGTCMSRVNSNFRPSKTVAPALQRCQTYMDIVLDGNSIYPWVE 180
Db 128 CGPLWAHQCGNQYATGTCSDVSPDFQFLTSFSDPAVQACPSLVDVVVVCDSNIIYPWEA 187
QY 181 VQHLINLLKFIYIGPGQIQVGVVQYGEDVVHFEHLNDYRSVKDVEAASHIEQRGGTET 240

```


Db 188 VKNFLVFKFTGLDIPGKTKTOVALIOYANEPRIIFNLNDEFTKEDMWQATSETRQGGDUT 247
Qy 241 RTAFGEFARSAFOK--GGRGAKKVMIVITDGHSDPDLEKVIQQRSDNVTRYAVA 298
Db 248 NTFRAIEFARDVAYSOTSGRGATKVMVVITDGHSDGSKLTVIQOCDDEILFFGTA 307
Qy 299 VLGYNNRRGINPETFLNEIKYIASDPDDKHFPNVTDAAALKDVIDDALGRIPFSLGNTKN 358
Db 308 VLGYLNRNALDKNLIKKEIKAIATASTPTERYFNVVADEAALLKAGVLGEQIFSEGTVOG 367
Qy 359 ETSFLEMSOTGESSHV--EDCVLGAVGAVDNGAVLKEISAGKVIPIRESYLKEPPE 416
Db 368 GDNFQMEMAQVGSADYAFQNDILMUGAVGAFDWSGLTVOETSHKPEVI-----FPK 418
Qy 417 EL-----KNHGAYLGVTTSVSSRQGRVYVAGAPRFNHTGKVIIFTMNNRSITIHQ 469
Db 419 QAPDQVLODRNHSFLGYSV-AAISTEDGVHFVAGAPRANYTQGVLYSVNKGQNVTVIQ 477
Qy 470 AMRGQIGSYFGEITSDIDGVTVDVLLVGAPMYFNE--GRERGKYVVELRQNRFPVN 528
Db 478 SHRGDQIGSYFVGLGSDVDKDTITDVLVGAPTYMNDLKKEGKVYILFTIKGILNOH 537
Qy 529 GTLKDSHSYQNAFSGSIASVRDLNODSYNDVVVGAPLEDNHAGAIYIFHGRGSIKTP 588
Db 538 QFLEGPETGNARFGSAIAALSIDINDGDFNDVIGSPVENSGAVIYINGHGTIRTKY 597
Qy 589 KORITASELA--TGLOFYGSIHQDLNEDGLIDLAVGALGNVILMRPVVQINASLH 646
Db 598 SQKILGNGAFRRHLQFFGSLDGYGLNGDSITDVSIGALGQVILQWSGIADVAIEAL 657
Qy 647 FEPKINIHRDCKRGRDATCLAAFLCTPIFLAHPQTITVIGIRYNATMD-----ERRY 702
Db 658 FTPDKITLLKDAK-----ITLKLCPRAEF--RPAGQNNQVAILFNMTLADGHSRV 708
Qy 703 TPAHLDEGDRTRNAVLLSSQELCERINFVLDTADYKVPVTSVEYSLBDPDHGP 762
Db 709 TSGVTFRENERFLQKMMVNEVQK--CSEHISIQPDSVDVNFELDURVDSLENPGTSPA 767
Qy 763 LDPGWETTLRV--SVFPNMGNEDEHCVDPDLVLDARSDDLPTAMEYQORVLKPKAQDCSAYT 821
Db 768 L-BAYSETVKVFSIPFYKEGSGDGCISDLILDVQ--QLP-----804
Qy 822 LSPDFTVPIETSRQVAVPATLENGENAYSTVNLISQSANLQFASLQKEDSDGSIEC 881
Db 805 -AIQTQSFIVSNQKRLTFVILKNRGEISAYNTWLAERSENLFASFMPVD---GTEV 860
Qy 882 VNERRLQKV--CNVSPYFPFRAKAVARLDSFSGSIFLHLEIELAAGSDSNERDSTK 940
Db 861 TCEVGSQKSVTCDVGPALKSQOQVTFINFDENLQNLQNAINFQAFESQEQ--TNK 918
Qy 941 EDNVAPLPHLKYBADVLFTRSSLSLSHYEVLK-----SSLERVDGIGPPFSCIFRIQ--NL 995
Db 919 ADNSVSLTIPLLYDAELHLRSTNINFEISSDENAPSIVKSVEDIGKFP--IFSLKVTA 976
Qy 996 GLPFIHGMKMTIPIATRSNRLKLRLDPLTDEA--NTSCNIGNSNTEYRPTV-----1048
Db 977 GSAPVSAVLTPIHPQYTKENPLLYLTGTQTOAGDISC-----TAEINPLKPHTAPS 1031
Qy 1049 ----EEDLRAPQLNHSNDVVSINCNIR--LVPNQEIINFHLGNILRSILKALKYKSMKI 1103
Db 1032 VSFKNENFRHTKELDCITSCSNITCWLKDLHMKAEYFINVTRVWNRTEFAASTFQT--V 1089
Qy 1104 MVNAAIQOHSPITFREDDPSROIETFEISK--QEDWQVPIWIVGSLVGLGLLLALLVLA 1162
Db 1090 QLTAFAAEDITHNPQLFTEENAVTIPLMIMKPTKEABVPTGVIITGSIAGILLALLAMTAG 1149
Qy 1163 LRKLGFRESARRRREPGLDP 1182
Db 1150 LWKLGFPK--ROYKMGQNP 1167

RESULT 6

ITA2_BOVIN
ID ITA2_BOVIN STANDARD; PRT; 1170 AA.
AC P53710;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).
GN ITGA2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94193647; PubMed=7511592;
RA Kamata T., Puzon W., Takada Y.;
RT "Identification of putative ligand binding sites within I domain of
integrin alpha 2 beta 1 (VLA-2 CD49b/CD29).";
RL J. Biol. Chem. 269:9659-9663(1994).
CC -!- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,
COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT
RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-P-P-G-E-R IN
COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER
CELLS TO COLLAGENS. MODULATION OF COLLAGEN AND COLLAGENASE GENE
EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
EXTRACELLULAR MATRIX.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR ENBL; L25886; AAB59255.1; -.
DR PIR; I45914; I45914.
DR HSSP; P17301; LAOX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP_3.
DR Pfam; PF00357; integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; VWFA; 1.
DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
Platelet; Signal; Repeat; Polymorphism; Calcium; Magnesium.
KW Platelet; Signal; Repeat; Polymorphism; Calcium; Magnesium.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1170 INTEGRIN ALPHA-2.
FT DOMAIN 19 1121 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1122 1143 POTENTIAL.
FT DOMAIN 1144 1170 CYTOPLASMIC (POTENTIAL).
FT REPEAT 34 92 FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT DOMAIN 177 367 VWFA.
FT REPEAT ? ? FG-GAP 3.
FT REPEAT 423 475 FG-GAP 4.
FT REPEAT 477 538 FG-GAP 5.
FT REPEAT 540 599 FG-GAP 6.
FT REPEAT 604 656 FG-GAP 7.


```

FT CA_BIND 498 496 POTENTIAL.
FT CA_BIND 552 560 POTENTIAL.
FT CA_BIND 616 624 POTENTIAL.
FT SITE 472 474 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1146 1150 GFPR MOTIF.
FT DISULFID 72 81 BY SIMILARITY.
FT DISULFID 669 726 BY SIMILARITY.
FT DISULFID 778 784 BY SIMILARITY.
FT DISULFID 854 865 BY SIMILARITY.
FT DISULFID 1008 1039 BY SIMILARITY.
FT DISULFID 1044 1049 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1063 1063 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 580 580 G -> V.
FT VARIANT 588 588 R -> K.
FT VARIANT 725 725 R -> S.
SQ SEQUENCE 1170 AA; 128929 MW; EECFPLCSF2448PB1 CRC64;

Query Match 30.6%; Score 1901.5; DB 1; Length 1170;
Best Local Similarity 36.3%; Pred. No. 2.5e-115;
Matches 437; Conservative 238; Mismatches 437; Indels 91; Gaps 30;

QY 23 FNMTRKPRVPGSTAFPGYVQOHDISGNKWLVGAPLEINGYQKTDGVYKCPV--IH 80
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 19 YNVLGPKIRSPGSEQFGYAVQFPIPKGNLLVGSFWSGPPKRNMGDVYKCPVDLST 78
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 81 GNCTKLNIGRVT--LSNVSEKDNMLGLSLATNPKNDSFLACSPWSHECGSSYYITGMC 139
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 79 TTCEKXNLNQTSTMSNVTEMTNMSLGLTLTRNVGTGGFLTCGPLWAQCGSQYYITGVC 138
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 140 SEVNSNPFESKVAALQCQTYMDIVIVLDGNSIYIPVVEVQHFLINILKFPYGPQOI 199
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 139 SDVSPDFQLRTSFAPAVQTCPSFIIDVVVVCDSNIIYFPWDVAKNFLEKFPVGLDIGPTKT 198
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 200 QGVVQVQGEDVVHEPLNDYRSVKDWEAASHIEGGGTETRTAFGIFPAREAQ--KG 257
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 199 QMGLLOYANPRVVENILNFKSKDEMIKATSTQYGGDLNFTFKAIQIARDTAYSTAG 258
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 258 GRGKAKKVMIVITDGESHSDPLEKVIQOESERNTRYAVAVLGYNNRRGINPETFLNEI 317
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 259 GRPGATKVMVVVTDGESHGSKLKAVIDQCNKDNLRFGLVGLYNNALDITKNLKEI 318
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 318 KYIASDPDDKHEFNVTDEAALKDIDVALGDRIFSLGKINKNETSGLEMSQTGFSHVVE 377
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 319 KAIASIPTEHFNFVNSDEADLEKAGTIGEQIFSGTGVQGDNFQMEMSQVGFSAEYSP 378
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 378 DG--VLLGAVGADWNGAVLKETSAGKVIPLRESYLKEFPPEL---KNHGAVLGYVTNSV 432
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 379 QNNILMLGAVGADWSGTWQKTPGHIL-----FSKAQFEQILQDRNHSVLGVSAS- 432
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 433 VSSRGRRVYVAGAPRNFHTGKVLFTMHNNRSLTIHQAMRGOQIGSYFGSEITTSVDIDGD 492
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 433 ISTGNSHFVAGAPRANVTGQIVLYSVNENGNVTVIQSRGQDQIGSYFGSVLCAVDVNDK 492
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 493 GYTDVLLVGAAPMYNE--GRERKQVYVELRQNRVYNTGLKDSHYQNAFGSSIASVVD 551
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 493 TIITDVLVGAAPMYNDLKEGRRVYLTITKILNWHQFLEGNGLENARFGSAIAALSD 552
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 552 LNQDSYNDVVGAPLEDNHAGAIYIPHGFRGSLTKPKORITASELA--TGQYFGCSIH 609
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 553 INMDGFNDVIGSPLENQSGAVIYINGHEGMIRLRYQKILGSDRAFSSHLQYFGRSLD 612
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

QY 610 GOLDNEDGLIDLAYGALGNVILMSRPVQINASLHFPEPSKINIFHRDCKRSGRDATCL 669
   :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:

```

RESULT 7

```

ITR2 HUMAN
ID ITR2_HUMAN STANDARD; PRT; 1181 AA.
AC P17301;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
GN ITR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-44.
RC TISSUE=Endothelial cells;
RX MEDLINE=89308879; PubMed=2545729;
RA Takada Y., Hemler M.E.;
RT "The primary structure of the VLA-2/collagen receptor alpha 2 subunit
RT (platelet GPIa): homology to other integrins and the presence of a
RT possible collagen-binding domain.";
RL J. Cell Biol. 109:397-407(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA Rieder M.J., Armet T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.

```

CC RX MEDLINE=98019223; PubMed=9353312;
RA Emsley J., King S.L., Bergelson J.M., Liddington R.C.;
RT "Crystal structure of the I domain from integrin alpha2beta1";
RL J. Biol. Chem. 272:28512-28517(1997).
RN [4]
RP VARIANT HPA-5 (BR).
RX MEDLINE=94043762; PubMed=7901236;
RA Santoso S., Kalb R., Walke M., Kiefel V., Mueller-Eckhardt C.,
RA Newman P.J.;
RT "The human platelet alloantigens Br(a) and Br(b) are associated with a
RT single amino acid polymorphism on glycoprotein Ia (integrin subunit
RT alpha 2)";
RL J. Clin. Invest. 92:2427-2432(1993).
RN [5]
RP VARIANT GLU-534.
RX MEDLINE=20206009; PubMed=10744142;
RA Kroll H., Gardemann A., Fechter A., Haberbosch W., Santoso S.;
RT "The impact of the glycoprotein Ia collagen receptor subunit A1648G
RT gene polymorphism on coronary artery disease and acute myocardial
RT infarction";
RL Thromb. Haemost. 83:392-396(2000).
CC -!- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,
CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT
CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-P-P-G-E-R IN
CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER
CC CELLS TO COLLAGEN, MODULATION OF COLLAGEN AND COLLAGENASE GENE
CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
CC EXTRACELLULAR MATRIX.
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1. Interacts with HPS5.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- POLYMORPHISM: Position 534 is associated with platelet-specific
CC alloantigen HPA-5 (BR). HPA-5A/BR(A) has Lys-534 and HPA-5B/BR(B)
CC has Glu-534. HPA-5B is involved in neonatal alloimmune
CC thrombocytopenia (NAIT or NATP). The K534E polymorphism may play a
CC role in coronary artery disease (CAD).
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -!- DATABASES: NAME=PROW; NOTE=CD guide CD49b entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49b.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17033; CAA34894.1; -.
CC EMBL; AF512556; AAM34795.1; -.
CC PIR; A33998; A33998.
CC PDB; 1A0X; 25-NOV-98.
CC PDB; 1DZ1; 02-AUG-01.
CC Genew; HGNC:6137; ITGA2.
CC MIM; 192974; -.
CC GO; GO:0008305; C:integrin complex; TAS.
CC GO; GO:0005866; C:plasma membrane; TAS.
CC GO; GO:0004895; F:cell adhesion receptor activity; TAS.
CC GO; GO:0005518; F:collagen binding; TAS.
CC GO; GO:0007596; P:blood coagulation; TAS.
CC GO; GO:0007160; P:cell-matrix adhesion; TAS.
CC GO; GO:0007397; P:histogenesis and organogenesis; TAS.
CC InterPro; IPR000413; Integrin_alpha.
CC InterPro; IPR002035; VWFA.
CC Pfam; PF01839; FG-GAP; 3_
CC Pfam; PF00357; integrin_A; 1.
CC Pfam; PF00092; vwa; 1.
CC SMART; SMC0191; Int_alpha; 5.
CC SMART; SMC0327; VWA; 1.
CC -----
CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC PROSITE; PS02344; VWFA; 1.
CC KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
CC KX Platelet; Signal; Repeat; Calcium; Magnesium; Polymorphism;
CC 3D-structure.
CC FT SIGNAL 1 29 INTEGRIN ALPHA-2.
CC FT CHAIN 30 1181 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 30 1132 POTENTIAL.
CC FT TRANSMEM 1133 1154 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 1155 1181 INTERACTION WITH HPS5.
CC FT REPEAT 45 103 FG-GAP 1.
CC FT REPEAT 2 2 FG-GAP 2.
CC FT DOMAIN 188 378 VWFA.
CC FT REPEAT 378 433 FG-GAP 3.
CC FT REPEAT 434 486 FG-GAP 4.
CC FT REPEAT 488 549 FG-GAP 5.
CC FT REPEAT 551 610 FG-GAP 6.
CC FT REPEAT 615 667 FG-GAP 7.
CC FT CA_BIND 499 507 POTENTIAL.
CC FT CA_BIND 563 571 POTENTIAL.
CC FT CA_BIND 627 635 POTENTIAL.
CC FT SITE 1157 1161 GPEFR MOTIF.
CC FT DISULFID 83 92 BY SIMILARITY.
CC FT DISULFID 680 737 BY SIMILARITY.
CC FT DISULFID 789 795 BY SIMILARITY.
CC FT DISULFID 865 876 BY SIMILARITY.
CC FT DISULFID 1019 1050 BY SIMILARITY.
CC FT DISULFID 1055 1060 BY SIMILARITY.
CC FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 343 343 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 432 432 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 460 460 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 475 475 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 699 699 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1057 1057 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 1081 1081 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT VARIANT 534 534 K -> E (IN ALLOANTIGEN HPA-5B;
CC gbsnp:1801106).
CC FT TURN 170 171 /FTid=VAR_003977.
CC FT STRAND 171 180
CC FT TURN 173 180
CC FT TURN 183 184
CC FT HELIX 188 199
CC FT TURN 200 201
CC FT STRAND 204 204
CC FT TURN 206 207
CC FT STRAND 209 216
CC FT STRAND 220 224
CC FT TURN 226 228
CC FT TURN 232 240
CC FT TURN 241 241
CC FT HELIX 252 262
CC FT TURN 263 264
CC FT HELIX 266 268
CC FT TURN 269 269
CC FT STRAND 275 282
CC FT HELIX 289 291
CC FT HELIX 292 301
CC FT TURN 302 303
CC FT STRAND 304 311
CC FT HELIX 313 317
CC FT TURN 318 319
CC FT HELIX 323 330
CC FT TURN 331 332
CC FT HELIX 337 340
CC FT STRAND 341 344
CC FT HELIX 347 353
CC FT HELIX 354 362
CC FT TURN 363 363
CC SQ SEQUENCE 1181 AA; 129295 MW; 7E1B7ED968A94070 CRC64;

[illegible]

```

QY 594 NLGLFPPHGMKKITIIPIATSGNRLLKLRDLFTLDEA-NTSCNIGWNGSTYRPTP----- 104
DQ 978 TTGSPVPSMATVITHIPQYTKKPKMLYLTGCVQDKAGDISCNADINPLKIGQTSVSVF 1037
QY 1048 VEEDLRAPOLNHSNDVWSINCIRLVPQCEINFHLG-----NLWLSLKALKYK 1059
DQ 1038 KSENFRTKELNCRATSCSNVTCWK-----DVHMKGEYFVNVTTRWNGTFASSTFQ 1090
QY 1100 SMKIMVNAALQORQSPHPIFREEDPSRQIEFEISK-QEDWQVPIIWIIVSGTLGLLLIAL 1158
DQ 1091 T--VQLTAAAEINTYNEPIVIEDNTVTIPLIMMKPDEKAEVPTGVIGSIAGILLLLA 1148
QY 1159 LVIALRLKLGFR 1170
DQ 1149 LVAILKLGFFK 1160

RESULT 8
ID _HUMAN
ID _ITAL HUMAN STANDARD; PRT; 1170 AA.
AC P20701; Q43746;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-L precursor (Leukocyte adhesion glycoprotein LFA-1
DE alpha chain) (Leukocyte function associated molecule 1, alpha chain)
DE (CD11a).
DE GN ITGAL OR CD11A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
[1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.
RP MEDLINE=89139587; PubMed=2537322;
RA Larson R.S., Corbi A.L., Berman L., Springer T.;
RT "Primary structure of the leukocyte function-associated molecule-1
RT alpha subunit: an integrin with an embedded domain defining a protein
RT superfamily.";
RT J. Cell Biol. 108:703-712(1989).
[2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RP MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Degliattis Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
RT Genomics 50:295-308(1999).
[3]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 153-335, AND REVISION TO 214.
RP MEDLINE=96036067; PubMed=7479767;
RA Qu A., Leahy D.J.;
RT "Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha
RT L beta 2) integrin.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:10277-10281(1995).
[4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 153-335.
RP MEDLINE=96398682; PubMed=8805579;
RA Qu A., Leahy D.J.;
RT "The role of the divalent cation in the structure of the I domain
RT from the CD11a/CD18 integrin.";
RL Structure 4:931-942(1996).
[5]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 153-334.
RP MEDLINE=99425288; PubMed=10493852;
RA Kallen J., Welzenbach K., Ramage P., Geyl D., Kriwacki R., Legge G.,
RA Cottens S., Weitz-Schmidt G., Hommel U.;
RT "Structural basis for LFA-1 inhibition upon lovastatin binding to the
RT CD11a I-domain.";
RT J. Mol. Biol. 292:1-9(1999).
RC -!- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,

```

CC	ICAM3 and ICAM4. IT IS INVOLVED IN A VARIETY OF IMMUNE PHENOMENA INCLUDING LEUCOCYTE-ENDOTHELIAL CELL INTERACTION, CYTOTOXIC T-CELL MEDIATED KILLING, AND ANTIBODY DEPENDENT KILLING BY GRANULOCYTES AND MONOCYTES.
CC	-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-L ASSOCIATES WITH BETA-2.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=2;
CC	Name=1;
CC	IsoId=P20701-1; Sequence=Displayed;
CC	Name=2;
CC	IsoId=P20701-2; Sequence=VSP_002738;
CC	Note=No experimental confirmation available;
CC	-1- TISSUE SPECIFICITY: LEUKOCYTES.
CC	-1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.
CC	-1- SIMILARITY: Contains 1 VWFA domain.
CC	-1- SIMILARITY: Contains 7 FG-GAP repeats.
CC	-1- DATABASE: NAME=PROW; NOTE=CD guide/prow/cd/cdlla.htm".
CC	WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cdlla.htm".
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; Y00796; CAA68747.1; --
DR	EMBL; AAC02310; AAC31672.1; --
DR	PIR; S03308; S03308.
DR	PDB; 1LFA; 29-JAN-96.
DR	PDB; 1ZON; 07-DEC-96.
DR	PDB; 1ZOO; 07-DEC-96.
DR	PDB; 1ZOP; 07-DEC-96.
DR	PDB; 1CQP; 07-AUG-00.
DR	PDB; 1DGO; 03-FEB-00.
DR	PDB; 1MJN; 28-JAN-03.
DR	PDB; 1MQ8; 14-JAN-03.
DR	PDB; 1MQ9; 14-JAN-03.
DR	PDB; 1MQA; 14-JAN-03.
DR	Genew; HGNC:6148; ITGAL.
DR	MIM; 153370; --
DR	GO; GO:0008305; C:integrin complex; TAS.
DR	GO; GO:0006928; P:cell motility; TAS.
DR	InterPro; IPR000413; Integrin_alpha.
DR	InterPro; IPR002035; VWF_A.
DR	Pfam; PF01839; FG-GAP; 3.
DR	Pfam; PF00357; integrin_A; 1.
DR	Pfam; PF00092; vwa; 1.
DR	PRINTS; PR01185; INTEGRINA.
DR	PRINTS; PR00453; VWFADOMAIN.
DR	SMART; SMO0191; Int_alpha; 4.
DR	SMART; SMO0327; VWA; 1.
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR	PROSITE; PS00234; VWFA; 1.
KW	Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW	Signal; 3D-structure; Magnesium; Calcium; Repeat;
KW	Alternative splicing.
FT	SIGNAL 1 25
FT	CHAIN 26 1170
FT	DOMAIN 26 1088
FT	TRANSMEM 1089 1112
FT	DOMAIN 1113 1170
FT	REPEAT 42 91
FT	REPEAT 92 149
FT	DOMAIN 170 349
FT	REPEAT ? ?
FT	REPEAT 401 455
FT	REPEAT 457 516
FT	INTEGRIN ALPHA-L.
FT	EXTRACELLULAR (POTENTIAL).
FT	POTENTIAL.
FT	CYTOPLASMIC (POTENTIAL).
FT	FG-GAP 1.
FT	FG-GAP 2.
FT	VWFA.
FT	FG-GAP 3.
FT	FG-GAP 4.
FT	FG-GAP 5.

QY 120 ACSPLMSHSCGSSYYTTCMCRVNSNFRSKTVA-PALQRC-QTYMDIVIVILDGSSNYP 177
Db 110 ACDPGLSRICDQNTYLSGLCYLFRQNLQPMQGRFGFQECIKGNVDLVFLDFGMSUQF 169
QY 178 --WVEVQHFILNLIKFYIGPOIQGVVVOYGEDVVEHFLNDYRSVKDVAASHIEQR 235
Db 170 DEFQKILDFMKDVMKK--LSNTSYQFAAVQFSTYKTEFDSDYVVKRQDPDALLKHVKHM 227
QY 236 GGTETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGHSDSDPLEKVIQOQSRDNT 293
Db 228 -LLNTNTGAINVATEVEFREELGARPDPATKVLIIITDGEATDSGNIDAA-----KDI 280
QY 294 YVAVAVLGYNRRGINPEFLNEIKYIASDPDKHFENVTDS-AALKOIVDALGRIPESL 352
Db 281 RYIIGI-GKHFTKSSQET----LHKFASKPASE-FVKILTDFEKLKDLFTLELQKIYVI 334
QY 353 EGTNNE--TSPGLENSQTFSSHVEDGVLLGAAGVAYDMNGAVLKETSAGKVIPLR---- 407
Db 335 EGTSKQDLTSFNMELSSSGISADLSRGAHVAVGAKOM-----AGGFLLDKADLQ 385
QY 408 -BSYLKEPPEELKNHGAYLGYTVTSVSSRQGRVYVAGAPRNNHTGKVLFTM-----HN 461
Db 386 DTFIGNPLETPEVRAGLYGYTVTLPSRQKTSLLASGAPRYQHMGVRLVLFQPOGGGHW 445
QY 462 NRSLLTIHQAMRQOQIGSYFGSEITSDVDIDGVDVLLVAGAPMYNEGRERKGVVYELR 521
Db 446 SQVQIHH---GTQIGSYFGGELCGVDVDQDGETELLIGALPLFYGEQR-GGRVIYQIR 500
QY 522 QNRFYVGTLDKSHSYQNAFQSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGR 581
Db 501 QLGFEVSELOQDPGYPLGRFGEAITALTDINGDGLVDVAVGAPLEEQ--GAVYIFNGRH 558
QY 582 GSILKTPQRITPASELATGLOVFGCSIHGOLDNEDGLDLAVGALGNVILMSVPUVQI 641
Db 559 GGLSPQQRLEGTVLQSGIOWFGRSIHGVKDLGGLADVAVGAESQMVLLSPVWDM 618
QY 642 NASLHFPSKINIFHRDCK-----RSGRDATCLAAFLCFTPIELAPHFQITTV-GIRY 693
Db 619 VTLMSFSFAEIPVHEVCSYSTSNKMKGVNIT-----LCFOIKSLYQFQGRUVANITY 673
QY 694 NATMDERYTPRAHDEGDRFTNRAVLSSQELCERINPHV-LDTADYVKPVTFSVEY 752
Db 674 TLQLDGHF--TRRRGLFPGRHRLRNIAVTTMS-CTDFSHPFVCVQDLISPINVLNF 731
QY 753 SLEDPDHGP---MLDDCGWPTLRYS-----VFWNGCDEDEHCVDPDLVDARSDLPT 801
Db 732 SLWEEGTPRQRAQKDIPIILRPSLHSETWEIPFNKCGDKKCEANL----- 781
QY 802 AMEYQVRVLRPAQCSAYTISFDTTVFIESTRQORVAVEATLENRGENAYSTVLNISQS 861
Db 782 -----RVSPSPARSALRLTAFAS-----LSVELSLNLEEDATVQVLDLHFP 824
QY 862 ANLQFASI-IKEDSDSGIEC--VNEERRLQKV--CNVSPFPFPAKAKVAPRL----- 910
Db 825 PGLSERKVEMLKPHSQIPVSCBELPEESRLLSALSCNVSPPIFKAGHSVALQMMFNTLV 884
QY 911 DSEFSKSLFLHLELAEAGSDNSRDESKEDNVAPLFLHLYRADVL----- 958
Db 885 NSSWGDSEVELH-----ANVTGNNESSDLLEDNSATTIPIILYINILIQOEDSTLYVS 938
QY 959 FTRSSLSH-----YEVKLNS-----LERYDGIGPPFSCIPRIQLNGLFPIHGMWM 1005
Db 939 FTPKPKIHQVKHMYQVRIQPSIHDBNIPTLEAVGVGPQ-----PSEG--- 982
QY 1006 KITPIATRSNRLKURDLFTDEANTSCNIGNSTEYRPTFVE-EDLRRAPQLNHSND 1064
Db 983 ----PITHQ-----WSVQME-PPVPECHVEDLERLPLDAAEPCLP 1015
QY 1065 VVSINCINRLVNOBINPHLLGNLM-----RSLKAIKYSMKIMVNAALQRPQHSFPI 1118
Db 1016 GALFPCPV--VPROBILVQVITGLTELVEIEIASSMFSI-CSSLSISFNSS--KHFF--- 1067
QY 1119 FREEDPSRQIEFESKQEDWQVPIWIVGSLTGLGLLLALLVLALRLKLGFF-RSARRRE 1177

Db 1068 YGSNASLAQVMKVDVVTEKQM-LYLYVLSGIGGLLLLLLFIYLYKGVFFRNLKRW 1126
QY 1178 PG 1179
Db 1127 AG 1128
RESULT 9
ITAD HUMAN
ID ITAD HUMAN STANDARD; PRT; 1162 AA.
AC Q13349; Q15575; Q15576;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-D precursor (Leukointegrin alpha D) (CD11d) (ADB2).
OS ITGAD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=96111956; PubMed=8777714;
RA Van der Vliet M., Le Trong H., Wood C.L., Moore P.F., St John T.,
Staunton D.E., Gallatin W.M.;
RT "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-
3.";
RL J. Biol. Chem. 275:8959-8969(2000).
RN [2]
RP IMMUNITY 3:683-690(1995).
RN [3]
RP SEQUENCE OF 1-235 FROM N.A.
RX MEDLINE=20187620; PubMed=10722744;
RA Noti J.D., Johnson A.K., Dillon J.D.;
RT "Structural and functional characterization of the leukocyte integrin
gene CD11d. Essential role of Sp1 and Sp3.";
RL J. Biol. Chem. 275:8959-8969(2000).
RN [4]
RP SEQUENCE OF 467-537; 571-602; 633-666; 788-834 AND 910-1125 FROM N.A.
RX MEDLINE=96257236; PubMed=8666289;
RA Wong D.A., Davis E.M., LeBeau M., Springer T.A.;
RT "Cloning and chromosomal localization of a novel gene encoding a human
beta 2-integrin alpha subunit.";
RL Gene 171:291-294 (1996).
RN [5]
RP INTERACTION WITH VCAM1.
RX MEDLINE=93059842; PubMed=9841932;
RA Grayson M.H., Van der Vliet M., Sterbinsky S.A., Michael Gallatin W.,
Hoffman P.A., Staunton D.E., Bochner B.S.;
RT "alpha2beta2 integrin is expressed on human eosinophils and functions
as an alternative ligand for vascular cell adhesion molecule 1
(VCAM-1).";
RL J. Exp. Med. 188:2187-2191(1998).
RN [6]
RP INTERACTION WITH VCAM1.
RX MEDLINE=93370002; PubMed=10438935;
RA Van der Vliet M., Crowe D.T., Hoekstra D., Vazeux R., Hoffman P.A.,
Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;
RT "The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a
binding interface between I domain and VCAM-1.";
RL J. Immunol. 163:1984-1990(1999).
CC -!- FUNCTION: INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND
VCAM1. MAY PLAY A ROLE IN THE ATHEROSCLEROTIC PROCESS SUCH AS
CLEARING LIPOPROTEINS FROM PLAQUES AND IN PHAGOCYTOSIS OF BLOOD-
BORNE PATHOGENS, PARTICULATE MATTER, AND SENESCENT ERYTHROCYTES
FROM THE BLOOD.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-D
ASSOCIATES WITH BETA-2.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED MODERATELY ON MYELOMONOCYTIC CELL
LINES AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON
TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES, FOAM CELLS WITHIN
ATHEROSCLEROTIC PLAQUES, AND ON SPLENIC RED PULP MACROPHAGES.

QY 883 --NEERLQKQVCNYSYPPFRKAKVARELDSEFS--KSIFLHLEIELAAGSDSNERDST 939
 Db 849 VTEDEGLSRSCSNVHIFHEGSGNGTFLVFDVSYKATLGRMLRASASSENKASS 908
 QY 940 KEDNVAPLRFHLKYEADVLFTSSLSHY-----EVKLSLSLRYDYGIPFSCIPRI 992
 Db 909 KA--TFQLELPVKYAVYIMISQSESTKYFNATSDKXKKEAHRV-----RV 955
 QY 993 QNLGFFPHGMMKTIPIATRSNRLKLRFLDEANTS--CNWGNSTYRTPVEE 1050
 Db 956 NNLQSORDU--AISINFWFVLLNG-----VAVWDVVMFAPSQSLPC-----VSEKPPQHS 1004
 QY 1051 D-----LRRAPOLNHSNDVVSINCRILVEN---QELNFHLLGNL---WLSLXALYK 1099
 Db 1005 DELTOISRSPMLDCSIADCLQRCB---VPSVSQBELDFTUKGNSLFGWVR--ETLQKK 1059
 QY 1100 SMKIMV--NAALQRFHSPFIREDPSRQIEPISKQEDWQVPIIIVGSTLGLLALL 1158
 Db 1060 VLIVSVAEITFTSVSYLPQGEAFMRQAMQEMVLEEDVYNA-IPIMGSSVGALLLAL 1118
 QY 1159 LVIALRLKLGFSARRRREPGLDTPK 1185
 Db 1119 ITATLYKLGF---KRHYKEMLEDKPE 1142

RESULT 10
 ITAX_HUMAN
 ID ITAX_HUMAN STANDARD; PRT; 1163 AA.
 AC F20702;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Integrin alpha-X precursor (Leukocyte adhesion glycoprotein p150,95
 DE alpha chain) (leukocyte adhesion receptor p150,95) (CD11c) (leu M5).
 GN ITGAX OR CD11C.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88166645; PubMed=3327687;
 RA Corbi A.L., Miller L.J., O'Connor K., Larson R.S., Springer T.A.;
 RT "cDNA cloning and complete primary structure of the alpha subunit of
 RT a leukocyte adhesion glycoprotein, p150,95.";
 RL EMO J. 6:4023-4028(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90153906; PubMed=2303426;
 RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;
 RT "Genomic structure of an integrin alpha subunit, the leukocyte
 RT p150,95 molecule.";
 RL J. Biol. Chem. 265:2782-2788(1990).
 RN [3]
 RP ERRATUM.
 RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;
 RL J. Biol. Chem. 265:12750-12751(1990).
 RN [4]
 RP SEQUENCE OF 20-43.
 RX MEDLINE=87167596; PubMed=3549901;
 RA Miller L.J., Wiebe M., Springer T.A.;
 RT "Purification and alpha subunit N-terminal sequences of human Mac-1
 RT and p150,95 leukocyte adhesion proteins.";
 RL J. Immunol. 138:2381-2383(1987).
 CC -!- FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT
 CC RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL
 CC INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY
 CC IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X
 CC ASSOCIATES WITH BETA-2.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND

CC GRANULOCYTES
 CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -!- SIMILARITY: Contains 1 VWFA domain.
 CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD11c entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd11c.htm".
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M81695; AAA59180.1; -;
 CC EMBL; Y00093; CAA68283.1; -;
 CC EMBL; M29165; -; NOT ANNOTATED CDS.
 CC EMBL; M29487; AAA51620.1; ALT SEQ.
 CC EMBL; M29482; AAA51620.1; JOINED.
 CC EMBL; M29483; AAA51620.1; JOINED.
 CC EMBL; M29484; AAA51620.1; JOINED.
 CC EMBL; M29485; AAA51620.1; JOINED.
 CC EMBL; M29486; AAA51620.1; JOINED.
 CC PIR; A36584; RWHLIC.
 CC PDB; IN3Y; 1B-FEB-03.
 CC Genew; HGNC:6152; ITGAX.
 CC MIM; 151510; -;
 CC GO; GO:0008305; C:integrin complex; TAS.
 CC GO; GO:0004895; F:cell adhesion receptor activity; TAS.
 CC GO; GO:0004872; F:receptor activity; TAS.
 CC GO; GO:0007155; P:cell adhesion; TAS.
 CC GO; GO:0007397; P:histogenesis and organogenesis; TAS.
 CC InterPro; IPR000413; Integrin_alpha.
 CC InterPro; IPR02035; VWF_A.
 CC Pfam; PF01839; FG-GAP; 3_
 CC Pfam; PF00357; integrin_A; 1.
 CC Pfam; PF00092; vwa; 1.
 CC PRINTS; PR01185; INTEGRINA.
 CC PRINTS; PR00453; VWFADOMAIN.
 CC SMART; SM00191; Int_alpha; 5.
 CC SMART; SM00327; VWF_1
 CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 CC PROSITE; PS50234; VWFA; 1.
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Magnesium; Calcium; Repeat; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 1163 INTEGRIN ALPHA-X.
 FT DOMAIN 20 1107 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1108 1128 POTENTIAL.
 FT DOMAIN 1129 1163 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 34 87 FG-GAP 1.
 FT REPEAT ? ? FG-GAP 2.
 FT DOMAIN 165 351 VWFA.
 FT REPEAT ? ? FG-GAP 3.
 FT REPEAT 402 453 FG-GAP 4.
 FT REPEAT 455 517 FG-GAP 5.
 FT REPEAT 518 576 FG-GAP 6.
 FT REPEAT 581 633 FG-GAP 7.
 FT CA_BIND 466 474 POTENTIAL.
 FT CA_BIND 530 538 POTENTIAL.
 FT CA_BIND 593 601 POTENTIAL.
 FT SITE 1131 1135 GFFKR MOTIF.
 FT DISULFD 69 76 BY SIMILARITY.
 FT DISULFD 108 126 BY SIMILARITY.
 FT DISULFD 555 712 BY SIMILARITY.
 FT DISULFD 771 777 BY SIMILARITY.
 FT DISULFD 848 863 BY SIMILARITY.
 FT DISULFD 998 1022 BY SIMILARITY.
 FT DISULFD 1027 1032 BY SIMILARITY.
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	89	89	N-LINKED (GLCNAC. . .)	(POTENTIAL).
CARBOHYD	392	392	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	697	697	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	735	735	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	899	899	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	939	939	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CARBOHYD	1050	1050	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT CONFLICT	490	490	G -> A (IN REF. 21)	
FT CONFLICT	756	756	G -> A (IN REF. 21)	
SQ SEQUENCE	1163 AA; 127885 MW; 6C4E19C3F62A473	CRG64;		

Query Match 18.2%; Score 1132.5; DB 1; Length 1163;
Best Local Similarity 28.4%; Pred. No. 2.1e-65;
Matches 354; Conservative 219; Mismatches 471; Indels 201; Gaps 48

Qy	7	LVAWALSLWPGFTDTNMDTRKPRVPGSR	TAF-----	FGYTVQQHDSGNKWL	VYGA 60	
Db	10	LFTALATSL-----	GFNLDTDE-----	LTAFRVDSAGFGDS	VWQY---ANSVVVYGA 53	
Qy	61	PLEINGYQKTGVYKCPVIHGNCYKLANLGRVTL	SNVSEKDNMRGLGSLATNPRD	NSFLA 120		
Db	54	PQKLTAAQNTGGLYQCGYSTGACBPGL-----	QVPEAVNMSLGLSLASTT	SPSULLA 107		
Qy	121	CSPLWSHBCSGSYVTGMC	SRVNSFRFSKTVAPALQRC-QTYMDIVIVLDC	SNIYP-- 177		
Db	108	CGPTVHHCEGRMMVLTGLCFL	LPGT-QLTQRLPYSRQEC	PROBODIVELIDGSG	ISSRN 166	
Qy	178	WVFQHFILINTLKKPYIGPGQIQGVQVQYGEDV	VHFEHLNDYRSVKD	VVEASHIEQ-RG 236		
Db	167	FATWMNFVRAVISOFQ--RPSTQFSLMQFSNKF	QTHFTFEFRRTSNPLSLAS	VHQIQG 224		
Qy	237	GTEITAFGIEFARSEAPQKGRKAKKMWIVITD	GESH-DSPDLKEVIOQSERDN	VRY 295		
Db	225	FYYTATALQNVHRLFHASYGARRDATKILIVITD	CKEKGSLDYKQVIMADA	GIIRY 284		
Qy	296	AVAV-LGYNNRRGINPEFTFLNEIKYIASDP	DDKHFNVTD	EAALKDIDVADG	RIFSLG 354	
Db	285	AIGVGLAFQNRNS-----	WKLNDIASKPSOEHI	FKVEDFDALXDIQNLKEK	IFATG 338	
Qy	355	T-NKNETSFGLEMSQTGFSSHVVEGVL	LGVAGYDNGAVL	KETSACKVLP-LRESYL 411		
Db	339	TETSSSFELEMAQEGSAVFTPDG	PVLGAVGSFTWGGAP-----	LYPNMSPTFI 391		
Qy	412	KEPPEELKNHGAYLGYTVTSVVS	SRQGVYVAGAPRNFHTK	GVILFTMHNRS	LTIHQAM 471	
Db	392	NMSQENVMDRDSYLGYS--TELALWKG	VSGLVLAGPRYQHTG	KAVIFT-QVSRQWRMA	KAEV 449	
Qy	472	RQOIGSGYFGEITISVDIDG	VDTVLLVGPAMYNEGR	KRYVYELRQ--NRFVYNG 529		
Db	450	TGTQIGSYFGASLCSVDVDTD	GSVDTLVLIGAPHYEQTR-GGQV	SVCPPLGRWRKWCDA 508		
Qy	530	TLKDSHSYQNAFSGSSIASVRDL	NDNSYDVVVGAPLEDNI	AGATYIPHEFG-SILKTP 588		
Db	509	VLYGEQHPWGRFGAALTVL	GDVNGDKLTDVVI	GAPBEENRGAVLHGV	LGPISLSPH 568	
Qy	589	KQRTASBLATGLQYFGCSIH	QDLNEDGLIDLAVGAL	CNVIWMSRPVQV	INASLHFE 648	
Db	569	SQRTAGSOLSRLOFYGOALS	GGQLTDQGLVDLAVGARGQ	VILLITRPLVMV	GVSMQFI 628	
Qy	649	PSKINIFHRDCKRS-GRDAT	CLAAFLCF-----	TFIPLAPHFTTTGIRY	NATMDRER 701	
Db	629	PAEIPRSAFECREQVVS	QGTIVQSNICLYDKRS	KNLLGSRDLQSS-----	VILDLALDPCR 685	
Qy	702	YTPRAHLDGGDRFTNR	AVILSSGGELCERIN	FHVL---DTADYKVP	FTVSVEYSLDPP- 757	
Db	686	LSPRATFOETKNRSLSR	VRVLGL-KAHCE--NFNILLPS	CVBDSVTPIRLN	FTLVGVKP 742	
Qy	758	-----DHGPM	LDCGWPTTLRVSP	PWNGNEDEHCV	PDVLVDARS	DLPTAMEYCORVLAK 812
Db	743	LLAFNLREPLAAL	AQRAFTASLP	KEKNCADHIC-----		777
Qy	813	PAQDCSAYTSLSDTT	TVTFIESTRQVRV	AEATLEN	GENAYSTVLINIS	OSANIQFASLI-- 870

Db 783 TYSAMGLDT---LVVGGPQDNMNSVTLRNDGEDSYGTQVTVYYPVSGLSYRKDSASQNPILT 839
QY 869 ----LIQKEDSGSIECVNEERRLQKOVNVSYPFFRAKAKVAFRLDSEF-SKSIPLHL 923
Db 840 KKPWFVKPAESSSSSE---GHGALKSTWNINHPANSEVTFNFTVDVDSHSPGNKL 896
QY 924 EIELAGSDSNERDSTKEDNVAFLRHLKYEADVLETRSSLSHY-----EVLKNS 975
Db 897 LUKAVASENM--SSTHTKTKFOLELPVKYAIYIVTSDSSIRYLNFTASEMTSKVIQH 954
QY 976 LERYDIGP---PFSICF---RIQNLGLFPPIHGMKMITIPIATRSNGNRLKLRLDFLD 1028
Db 955 QYQFNLIQORSLPVSFVFPVQINNVTVD-HPQVI-----FSQ 993
QY 1029 EANTSNIGWNSTEYRTP---VEEDLRAPQLNHSNDVVSINCMIRLVNQEI-NFHL 1084
Db 994 NLSSACH---TEQSPHSPNPRDQLERTFVLNCSVAVCKRIQCDLPSPFNTQEIFNVTL 1048
QY 1085 LGNL---WLSLKL---ALYKSMKIMVNAALQRFHSPFIREDPSRQIEFEISKQEDW 1138
Db 1049 KGNLSFDWDIKSHGHLILVSVSTEILFN-----DSFALLPGQSYVRSKTETKVEPY 1101
QY 1139 QV---PIWIIVSGTLGLLILALVLAIRKLGFPR 1170
Db 1102 EVHNPVLIVGSSIGGLVLLALITAGLYKLGFPK 1135
RESULT 12
ITAM HUMAN
ID ITAM HUMAN STANDARD; PRT; 1152 AA.
AC P11215;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MOI)
DE (Neutrophil adherence receptor).
GN ITGAM OR CR3A OR CD11B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88315033; PubMed=2457584;
RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;
RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the integrins, von Willebrand factor and factor B.";
RL J. Biol. Chem. 263:12403-12411(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88190151; PubMed=2833753;
RA Arnaout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;
RT "Molecular cloning of the alpha subunit of human and guinea pig leukocyte adhesion glycoprotein MOI: chromosomal localization and homology to the alpha subunits of integrins.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=88257215; PubMed=2454931;
RA Arnaout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;
RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor MOI (complement receptor type 3).";
RL J. Cell Biol. 106:2153-2158(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93123748; PubMed=8419480;
RA Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;
RT "Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-integrin gene family demonstrate remarkable conservation of genomic organization and suggest early diversification during evolution.";

RL J. Immunol. 150:480-490(1993).
RN [5]
RP SEQUENCE OF 9-1153 FROM N.A.
RX MEDLINE=8909893; PubMed=2563162;
RA Hickette D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L., Roth G.J.;
RT "cDNA sequence for the alpha M subunit of the human neutrophil adherence receptor indicates homology to integrin alpha subunits.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).
RN [6]
RP SEQUENCE OF 1-9 FROM N.A.
RX MEDLINE=92073318; PubMed=1683702;
RA Shelley C.S., Arnaout M.A.;
RT "The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).
RN [7]
RP SEQUENCE OF 1-9 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=92144986; PubMed=1346576;
RA Pahl H.L., Rosmarin A.G., Tenen D.G.;
RT "Characterization of the myeloid-specific CD11b promoter.";
RL Blood 79:865-870(1992).
RN [8]
RP SEQUENCE OF 17-31.
RX MEDLINE=87076671; PubMed=3539202;
RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaout M.A.;
RT "N-terminal sequence of human leukocyte glycoprotein MOI: conservation across species and homology to platelet IIb/IIIa.";
RL Biochim. Biophys. Acta 874:368-371(1986).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.
RX MEDLINE=95171458; PubMed=7867070;
RA Lee J.O., Rieu P., Arnaout M.A., Liddington R.;
RT "Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18).";
RL Cell 80:631-638(1995).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.
RX MEDLINE=96363671; PubMed=8747460;
RA Lee J.O., Bankston L.A., Arnaout M.A., Liddington R.C.;
RT "Two conformations of the integrin A-domain (I-domain): a pathway for activation?";
RL Structure 3:1333-1340(1995).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.
RX MEDLINE=98362595; PubMed=9687375;
RA Baldwin E.T., Sarver R.W., Bryant G.L. Jr., Curry K.A., Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrichson R.L., Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E., Matchler V.F., Tomich C.S., Watenpaugh K.D., Wiley V.H.;
RT "Cation binding to the integrin CD11b I domain and activation model assessment.";
RL Structure 6:923-935(1998).
RN [12]
RP 3D-STRUCTURE MODELING OF 17-616.
RX MEDLINE=98226734; PubMed=9560195;
RA Oxvig C., Springer T.A.;
RT "Experimental support for a beta-propeller domain in integrin alpha subunits and a calcium binding site on its lower surface.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).
CC -!- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES. IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M ASSOCIATES WITH BETA-2.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND

Qy	194	IGBQIQGVVQVGVGDVWHEFHNDYRSVKDVVEASHIEQRCGTTTRTAFAGTEPARSEA	253
Db	227	EKCFECNFALVOYGAIVOTEFDLQESRDINASLAKVQSIVQKVEV--TKTASAMQHVLDNI	285
Qy	254	F--QKGRGAKKVMIVITDGESHDSF--DLEKVIQOSERDNTVRYAVVLGYNNRGINP	310
Db	286	FIPSRGSRKALKVMVLITDGLFGDPLNUTTVINSPKMGQVVRFAIGV---GDRFKNN	341
Qy	311	ETFLNBKIYASDPDDKHFFNFVDEAALDKIDVDAJGDRIFSLRGTNKNETSFGLSEMSQTG	370
Db	342	NTY--RELKLIASDPKEAHTFKVINYSALDGLLSKLQORIVHMBGTVDGALQY--QLAQTG	398
Qy	371	FSHHVVEDG--VLLGAVGADWN--GAVIKETSACKVTPLRSEYILKEPFEELKN--HCAYILG	426
Db	399	FSAQIILDGQVLGTVGAFNFGSGGALLYSTQNG---RGCFJNQTAKEDSRVTQVSYLLG	453
Qy	427	YTVTSVVSRQGRVYVAGAPRFNHTKGVILFTMHNNRSLTIHQAMRGQOIGSYFGSGEITS	486
Db	454	YSL--AVLHRAHGISYVAGAPRHKLRCAGVFLRKEDREEDAFVRRIEGQMGSYFGSVLCP	512
Qy	487	VDIDGVDTVLLVVGAPMVFNEGRGKVKVYVELRON--RFVYNGTLKDSHVSQONARFGS	544
Db	513	VDIDMGDTTDFLVAAPFFYHIRG--BEGRVVYVQVPQDQASFSLAHTLSGHPGLTNRFGF	571
Qy	545	SIASVRDLNODSYNDVVVGCAPLEDNEHA-----GAIYVPHGFRGSLTKTPKORITASELA	598
Db	572	AMAAVGDIQDKFTDVAIGAPLEGFAGAGDQASVGSYIINGHSGGLYDSEFSQOIRASSVA	631
Qy	599	TGHQYFGCSITHGQLDNEGLIDVLAGALGNVILMSRPVQVINASLHPFPS-----650	
Db	632	SGLHYFGMSVSGGLDFNGDGLADITVGSRDSAVVLSRSPVVDLTVSMITTPDALPMVFIG	691
Qy	651	--KINI--PHRDCKRGRDATCLAAFLCFPTPIFLAPHFOITTVGIRYNATWDERRYTPRAH	707
Db	692	KMDVNLCEFYDSSVASSEPLREMFNLFT-----VDVDVTQQRQLQ	733
Qy	708	LDE-----GGRFTNRVALLSSGOELCERINFHVLDTADVVPKPTVFSEVYSLED	756
Db	734	CEBSSGCGQCLRKWNGGSFLCEHFWLLSTIEELCEE-----DCFSNITIKVTEFQT	784
Qy	757	P-----DH-GPMLDDGNTTILRSVSPFWNGNEDEHCVPLDLVLDARSDLPTAMBYCORVLR	811
Db	785	SGGRBDYVNPDLTHYKEPSAIFOLPYKECKCKVFCIAETQL-----826	
Qy	812	KPAQDCSAYLTFDITVFIIESTRQVAVEATLENGENAYSVTLNISQANLQFASLIQ	871
Db	827	-----FTNISQOELVGVVTKY--VTWNISUTNSGEDSYMTNMAINTPRNLQFKK-IQ	875
Qy	872	KEBDSGSIECVNEERLRQKV--CNVSYPPFFRAKAKVAFRLDSEFSKIFLHHEIELAA	929
Db	876	KPVSP--DVQCDPDKPVASVLWNCKIGHPILK--RSSVNVSVTWQLEESVFPNRTADITVT	933
Qy	930	GSDSNRDSYKENDVAPLRHLKYEA-----DVLFF--TRSSSLSHVEVKLNSLSERYDGI	982
Db	934	ISNSNEKSLARE--TRSLOFRHAFI AVLRSPPSYMYNMTSQSPSDKHKEFFNFVHGENL---	988
Qy	983	GPFPSCIFRIQNLGLPPIHGMMMKIITPIATRSGNRLLLKLRDLTDEAMTSCMIGWNSTE	1042
Db	989	---FGAVFOLQ-----ICVPFKLOD--FOIVRVKNLTKTQDHTEC-----1023	
Qy	1043	YRPTPVEEDLRRAPQNLNH-----SNSDVVSYINCNTIBLVNPQNEINFHLLGNLWL	1090
Db	1024	---TQSGEPACGSDPPQHVKEHWSVVCATSNKENTVVAEIVSGVHTKOL-----L	1071
Qy	1091	RSLKALKY-----KSMKIMVNAALQROPHSPFIFREDEPSPQIEFISKQEDWQVPIW	1143
Db	1072	RDVSELPILGEISFNSKSLYEGINAEHNRHT-KITVIFLKEEETRSLP-----1116	
Qy	1144	IIYVSTGGLLLLALLVLAALRKLGFPSRRRRREPLDPT	1183
Db	1117	LIIGSSIGGLLVAVIIAILFKCGFKK--RKYOOLNLEST	1154

```

RESULT 15
ID ITAE HUMAN STANDARD; PRT; 1179 AA.
AC P38570; Q9NZU9;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-E precursor (Mucosal lymphocyte-1 antigen) (HML-1
DE antigen) (CD103 antigen) (Integrin alpha-IE1).
GN ITGAE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-38 AND 179-188.
RC TISSUE=Leukemia, and Lymphocytes;
RX MEDLINE=94164962; PubMed=8119947;
RA Shaw S.K., Cepek K.L., Murphy E.A., Russell G.J., Brenner M.B.,
RA Parker C.M.;
RT "Molecular cloning of the human mucosal lymphocyte integrin alpha E
RT subunit. Unusual structure and restricted RNA distribution.";
RL J. Biol. Chem. 269:6016-6025(1994).
RN [2]
RP REVISIONS TO 88-114.
RA Parker C.M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 53-1179 FROM N.A.
RC TISSUE=Fetal kidney;
RX MEDLINE=20138496; PubMed=10673275;
RA Touchman J.W., Anikster Y., Dietrich N.L., Maduro V.V., McDowell G.,
RA Sholevaruk V., Bouffard G.G., Beckstrom-Sternberg S.M., Gahl W.A.,
RA Green E.D.;
RT "The genomic region encompassing the nephropathic cystinosis gene
RT (CNS): complete sequencing of a 200-kb segment and discovery of a
RT novel gene within the common cystinosis-causing deletion.";
RL Genome Res. 10:165-173(2000).
RN [4]
RP MUTAGENESIS OF ASP-109 AND PHE-316.
RX MEDLINE=20400502; PubMed=10837471;
RA Higgins J.M.H., Cernadas M., Tan K., Irie A., Wang J.-H., Takada Y.,
RA Brenner M.B.;
RT "The role of alpha and beta chains in ligand recognition by beta 7
RT integrins";
RL J. Biol. Chem. 275:25652-25664(2000).
CC -!- FUNCTION: INTEGRIN ALPHA-E/BETA-7 IS A RECEPTOR FOR E-CADHERIN. IT
CC MEDIATES ADHESION OF INTRA-EPITHELIAL T-LYMPHOCYTES TO EPITHELIAL
CC CELL MONOLAYERS.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAINS LINKED BY A
CC DISULFIDE BOND. ALPHA-E ASSOCIATES WITH BETA-7.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED ON A SUBCLASS OF T-LYMPHOCYTES KNOWN
CC AS INTRA-EPITHELIAL LYMPHOCYTES WHICH ARE LOCATED BETWEEN MUCOSAL
CC EPITHELIAL CELLS.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD103 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd103.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; L25851; AAB59359.2; -.

```

```

DR EMBL; AF168787; AAF43107.1; -.
DR PIR; A53213; A53213.
DR HSP; F1215; IABX.
DR Genew; HGNC:6147; ITGAE.
DR MIM; 604682; -.
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0004895; F:cell adhesion receptor activity; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00131; Int_alpha; 3.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Polymorphism; Magnesium; Calcium.
FT SIGNAL 1 18
FT CHAIN 19 1179 INTEGRIN ALPHA-E.
FT CHAIN 19 177 INTEGRIN ALPHA-E LIGHT CHAIN.
FT CHAIN 179 1177 INTEGRIN ALPHA-E HEAVY CHAIN.
FT DOMAIN 19 1124 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1125 1147 POTENTIAL.
FT DOMAIN 1148 1179 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 181 198 GLU-RICH (ACIDIC).
FT REPEAT ? ? FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT REPEAT ? ? X-DOMAIN (EXTRA DOMAIN).
FT DOMAIN 145 199 VWFA.
FT DOMAIN 200 391 FG-GAP 3.
FT REPEAT 401 456 FG-GAP 4.
FT REPEAT 457 506 FG-GAP 5.
FT REPEAT 510 571 FG-GAP 6.
FT REPEAT 573 638 FG-GAP 7.
FT REPEAT 641 693 POTENTIAL.
FT CA_BIND 522 530 POTENTIAL.
FT CA_BIND 586 594 POTENTIAL.
FT CA_BIND 654 662 POTENTIAL.
FT SITE 1150 1154 GFFKR MOTIF.
FT DISULFID 70 79 BY SIMILARITY.
FT DISULFID 126 159 BY SIMILARITY.
FT DISULFID 706 762 BY SIMILARITY.
FT DISULFID 823 829 BY SIMILARITY.
FT DISULFID 893 907 BY SIMILARITY.
FT DISULFID 1008 1033 BY SIMILARITY.
FT DISULFID 1041 1057 BY SIMILARITY.
FT CARBOHYD 49 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 726 726 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 934 934 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 954 954 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1065 1065 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1096 1096 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 360 D -> E.
FT VARIANT 1041 C -> S.
FT MUTAGEN 208 /FTid=VAR_008885.
FT MUTAGEN 316 D->A: LOSS OF E-CADHERIN BINDING.
FT CONFLICT 477 F->A: LOSS OF E-CADHERIN BINDING.
FT CONFLICT 482 Q -> R (IN REF. 3).
FT CONFLICT 950 R -> W (IN REF. 3).
FT CONFLICT 1019 A -> V (IN REF. 3).
SQ SEQUENCE 1179 AA; 130088 MW; E558902EDFD9D5E1 CRC64;

```

Query Match
Best Local Similarity

14.9%;
Score 924.5; DB 1; Length 1179;
Pred. No. 6.9e-52;

Matches	338; Conservative	206; Mismatches	482; Indels	259; Gaps	52;		
QY	23 FNMTRKPRVIFGSR	TAFFGYTVOQHDI	SGNK-WLVV	GAPLETNGYKGTGDVYKCPVI	ING 81		
Db	19 FNVDAVPWLITPKG	APFVLSLLHQDPST	NTQTL	LVTS-RTKTPGPHRCS	IVQD 75		
QY	82 NCTKMLGRVTL	SNVSRKDNMRGLG	LSATPKNS	FLACPLW-SHECGSYTYTGM	138		
Db	76 EILCHPVEHVPIK	GRHR-----GVT	VVRS--HEGV	LICIVLVRPHSLSS	E-LTGT 125		
QY	139 CS-----	RVNSNFRESK-----	TVAPAL	QRC-----	159		
Db	126 CSLLGPDLRPAQ	ANFFDENLLDP	ARVDTGDCV	SNKEGGEDDVNTAR	ORRALEKEEE 185		
QY	160 -----	QTYMDIVIVLDG	NSNYP--WVEV	QHFLINILKKFYIG	QIQGVVQ 205		
Db	186 EDKEEEDDEE	EAGTEIAIILDG	SGIDPPDFQ	AKDFISNMNRNFEYK	CEFCNFALVQ 245		
QY	206 YGEDVVEHFL	NDYRSVKDVVEA	ASHIEOR	GGTERTAFGIEFAR	SEAF--QKGRKGAK 263		
Db	246 YGVIQTEFL	RDSQVWMA	SLARVQNI	TQVGSV-TKTASAM	QHVULDSIFTS	SHGSRKAS 304	
QY	264 KWMIVITDG	ESHDP-DLEKVI	QOBERDN	VTRYAVAVILGY	YNRRGINPETF-----	LNEI 317	
Db	305 KVMVLTDG	GIPEDPLNLT	VINSPMQ	QVERFALGV-----	GEERFKSARTAREL 354		
QY	318 KYTASPD	KHPNVTEA	ALKDIDV	ALGDRI	FSLEGINKNETS	FGLEMSQTGESSHVE 377	
Db	355 NLITASPD	ETHAFKVTNY	WALDGL	SKURYNII	SMEGTVGDALHY--	QLAQIGFSAQILD 412	
QY	378 D-GVILLG	AVGADWN-GAVL	KETSAGK	VIPIRESYLKE	FFBELKNHGAYL	GYTVTSVSS 435	
Db	413 ERQVILLG	AVGADFWSG	ALLYDTR	SPRGREFLNQ	TAAAAADA	EAAQY-SYLGAV-AVLHK 470	
QY	436 RQRVTVAG	APRNTHTK	VILFTM	MNNRSLTHQ	AMRQQOIGSY	FGSEBITSV	VDIDGQVT 495
Db	471 TCSLSV	VAGAPQYKH	GA	VFEL-QKEG	REASFLPVL	EGEQMGSY	FGSELCPVDIDMDGST 529
QY	496 DVLVVG	AMPYENEG	RGRGVVY	VEL--RQNR	FVYNGTLK	DSHSYQNA	RFSGSSIASVRDLN 553
Db	530 DFLVAP	PFVHVH-G	EGRVYV	IRSEQ	GSFLARIL	SHGPGFTN	ARFGFAMAAGDLS 589
QY	554 QDSYND	VVVGAPLE---	DNHA-GAIY	IHFGRS	ILKTPK	QRTASE	LATGLQYFGCS 607
Db	589 QDKLTD	VAIGAPLE	FGADG	ASFGSV	YIYNGH	WDGLSASP	SQIRASTAVAPGLQYFGMS 648
QY	608 IHQOLD	INEGLIDL	AVGALGN	AVILWS	RPVQIN	ASLHFB	PSKINIPHRDCKESGRDAT 667
Db	649 MAGGFI	DGGLADIT	VGTVGL	GVVRGR	VPVRLK	VSMAFT	PSALPIGFNG----- 699
QY	668 CLAAFL	CTFTPI	FLAPH	FTTVGIR---	YNATMD	ERRYTP	RAHLD-----E 710
Db	700 VVMVRL	CFE--ISSV	TASEGL	REALN	FTLDV	DVGKQRR	RLQCSVRSCLGLREWS 756
QY	711 GGRFTN	RAVILSS	GGELCER	INFH	VDIADY	VKVF	TSVYSLEDP----DH-GPMLDD 765
Db	757 SCSQ	CEDELL	MPTE	GELCEE-----	DCFS	NASVKVS	YQLQTPREGQTDHPQPIILDR 807
QY	766 GWPTLR	VSVPFW	NGNEDE	HCVP	DILV	DARS	DLPTAMEYQORVLRKPAODCSAYTILSFD 825
Db	808 YTEP	FAIFQ	LPYEK	ACKNKL	FCVAEL	QL-----	ATTVSQQ 842
QY	826 TTVFTI	ESTRQ	VAEAT	LEN	GENAY	STVLN	ISOSANLQFASLIQKEDSDGSECVNEE 885
Db	843 E---	LVVGLT	KELTLN	INTNS	GEDSYNT	SMALN	PRNLQKRM-QKPPSP-NIQCDPQ 897
QY	886 RRLQ	KQV--CNV	SPPFR--	AKAV	APFLD	SEFSK	IFLHLELELAAGSDSNERDSTK 940
Db	898 PVASV	ILMCR	IGHPV	LKSSA	HSV	VWQLE-----	ENAFNRTADITVTVINSNERSLA 953
QY	941 EDNVA	PLRFL	KYEA-----	DVL	FTRS--SSL	SHYEVK	LNLSLERYDYGIPFPSCIFRIQN 994
Db	954 NE--	THTLQ	FRHGF	VA	LKSPS	IMVNT	GQGLSHHKE----- 988

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 13:03:59 ; Search time 214.039 Seconds
(without alignments)
1782.414 Million cell updates/sec

Title: US-09-980-403-2

Perfect score: 6224

Sequence: 1 MDLPRGLVVAWALSLWPGFT.....FRSARRRRPGLDTPKVLK 1188

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6196	99.6	1188	15	US-10-291-265-338 Sequence 338, App
2	6192	99.5	1188	15	US-10-291-265-810 Sequence 810, App
3	6188.5	99.4	1189	10	US-09-984-130-35 Sequence 35, Appl
4	6188.5	99.4	1189	10	US-09-836-353A-35 Sequence 35, Appl
5	6188.5	99.4	1189	12	US-10-262-839-4 Sequence 4, Appl
6	5780	92.9	1120	12	US-10-262-839-6 Sequence 6, Appl
7	5383	86.5	1034	10	US-09-984-130-43 Sequence 43, Appl
8	5383	86.5	1034	10	US-09-836-353A-43 Sequence 43, Appl
9	3621	58.2	707	9	US-09-764-870-313 Sequence 313, App
10	3621	58.2	707	14	US-10-125-540-313 Sequence 313, App
11	3182	51.1	688	10	US-09-866-050A-624 Sequence 624, App
12	3113	50.0	696	10	US-09-866-050A-501 Sequence 501, App
13	2518	40.5	545	10	US-09-866-050A-500 Sequence 500, App
14	2412.5	38.8	1167	16	US-10-741-601-531 Sequence 531, App
15	2412.5	38.8	1177	16	US-10-741-601-532 Sequence 532, App

16	2200	35.3	437	15	US-10-108-260A-3386 Sequence 3386, App
17	2166.5	34.8	1179	12	US-09-918-715-250 Sequence 250, App
18	2159	34.7	1151	10	US-09-984-130-103 Sequence 103, App
19	2159	34.7	1151	10	US-09-836-353A-103 Sequence 103, App
20	2149.5	34.5	1180	12	US-09-918-715-307 Sequence 307, App
21	1863	29.9	1181	12	US-10-211-462-187 Sequence 187, App
22	1863	29.9	1181	14	US-10-160-354-2 Sequence 2, Appli
23	1863	29.9	1181	15	US-10-295-027-1286 Sequence 1286, App
24	1809	29.1	1147	12	US-10-336-603A-42 Sequence 42, Appl
25	1194.5	19.2	1161	9	US-09-350-259-53 Sequence 53, Appl
26	1194.5	19.2	1161	10	US-09-891-943-53 Sequence 53, Appl
27	1184.5	19.0	1161	9	US-09-350-259-55 Sequence 55, Appl
28	1184.5	19.0	1161	10	US-09-891-943-55 Sequence 55, Appl
29	1181.5	19.0	1155	9	US-09-350-259-46 Sequence 46, Appl
30	1181.5	19.0	1155	10	US-09-891-943-46 Sequence 46, Appl
31	1167.5	18.8	1151	9	US-09-350-259-37 Sequence 37, Appl
32	1167.5	18.8	1151	10	US-09-891-943-37 Sequence 37, Appl
33	1146.5	18.4	1170	9	US-09-945-265-2 Sequence 2, Appli
34	1146.5	18.4	1170	12	US-10-261-164-1 Sequence 1, Appli
35	1142.5	18.4	1161	9	US-09-350-259-99 Sequence 99, Appl
36	1142.5	18.4	1161	10	US-09-891-943-99 Sequence 99, Appl
37	1139	18.3	1161	9	US-09-350-259-2 Sequence 2, Appli
38	1139	18.3	1161	10	US-09-891-943-2 Sequence 2, Appli
39	1136	18.3	1223	16	US-10-408-765A-295 Sequence 295, App
40	1132.5	18.2	1163	15	US-10-116-275-204 Sequence 204, App
41	1103.5	17.7	1163	9	US-09-350-259-4 Sequence 4, Appli
42	1103.5	17.7	1163	10	US-09-891-943-4 Sequence 4, Appli
43	1094	17.6	1153	9	US-09-350-259-3 Sequence 3, Appli
44	1094	17.6	1153	10	US-09-902-481A-1 Sequence 1, Appli
45	1094	17.6	1153	10	US-09-891-943-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-291-265-338

Query Match	99.6%	Score 6196;	DB 15;	Length 1188;
Best local Similarity	99.7%	Pred. No. 0;		
Matches 1184;	Conservative	0;	Mismatches	4;
			Indels	0;
			Gaps	0;
Qy	1	MDLPRGLVVAWALSLWPGFTDTFNMTRKPRVPGSRTAFPGYTVQOHDISGNKWLIVGA	60	
Db	1	MDLPRGLVVAWALSLWPGFTDTFNMTRKPRVPGSRTAFPGYTVQOHDISGNKWLIVGA	60	
Qy	61	PLETNGYQKTGDYKCPVHGNCCTKLNGLRVTLISNYSERKDNVRLGLSLATNPDKNSFLA	120	
Db	61	PLETNGYQKTGDYKCPVHGNCCTKLNGLRVTLISNYSERKDNVRLGLSLATNPDKNSFLA	120	

```
QY 121 CSPLWSHECGSSYYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSNYSYPWVE 180
Db 121 CSPLWSHECGSSYYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSNYSYPWVE 180
QY 181 VOHFLINILKKFYIGPGQIQGVVQYGEDVHVEHFLNDYRSVKVDVVEAASHIEQGGTET 240
Db 181 VOHFLINILKKFYIGPGQIQGVVQYGEDVHVEHFLNDYRSVKVDVVEAASHIEQGGTET 240
QY 241 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHSDSPDLEKVIQOQSERDNVTRYAVAVL 300
Db 241 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHSDSPDLEKVIQOQSERDNVTRYAVAVL 300
QY 301 GYNNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDIDVALGDRIFSLEGTNKNET 360
Db 301 GYNNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDIDVALGDRIFSLEGTNKNET 360
QY 361 SFGLEMSOTGFSSSHVVVEDGVLGAVGYDNGAVILKETSAKVIPLESYLYKEFPEELKN 420
Db 361 SFGLEMSOTGFSSSHVVVEDGVLGAVGYDNGAVILKETSAKVIPLESYLYKEFPEELKN 420
QY 421 HGAYLGYTVTSVVSRRQGRVYVAGAPRNFHTGKVLFTMHNRSITIHQAMRGOQIGSYF 480
Db 421 HGAYLGYTVTSVVSRRQGRVYVAGAPRNFHTGKVLFTMHNRSITIHQAMRGOQIGSYF 480
QY 481 GSEITSVSDIDGCVTDVLLVGAPMYFNEGRGKVVYVELRQNRVYNGTLKDSHYQNA 540
Db 481 GSEITSVSDIDGCVTDVLLVGAPMYFNEGRGKVVYVELRQNRVYNGTLKDSHYQNA 540
QY 541 RFGSSIASVRDLNQSDYNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQRTITASELATG 600
Db 541 RFGSSIASVRDLNQSDYNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQRTITASELATG 600
QY 601 LQYFGCSIHGQJDLNEDGLIDLAVGALGNVILWSRPVVQINASHLHFEPKINIFHRDCK 660
Db 601 LQYFGCSIHGQJDLNEDGLIDLAVGALGNVILWSRPVVQINASHLHFEPKINIFHRDCK 660
QY 661 RSGRATCLAAFLCPTPIPLAHPQTTTIGIRYNATMBERRYTPRAHIDEGGDRTNRAV 720
Db 661 RSGRATCLAAFLCPTPIPLAHPQTTTIGIRYNATMBERRYTPRAHIDEGGDRTNRAV 720
QY 721 LLSGQELCERINFHVLDTADYVKPVTFSVEYSLEDPDHGMPLDDGWPITLRVSVFPWNG 780
Db 721 LLSGQELCERINFHVLDTADYVKPVTFSVEYSLEDPDHGMPLDDGWPITLRVSVFPWNG 780
QY 781 CNEDEHCVDPDLVDARSDLPTAMEYQCVLRKPAQDCSAYTILSFDTTVPFIIESTRQAV 840
Db 781 CNEDEHCVDPDLVDARSDLPTAMEYQCVLRKPAQDCSAYTILSFDTTVPFIIESTRQAV 840
QY 841 EATLENRGENAYSTVLNITSOSANLOPASLIQKEDSDGSTECVNEERRLOKVCNVSYPFF 900
Db 841 EATLENRGENAYSTVLNITSOSANLOPASLIQKEDSDGSTECVNEERRLOKVCNVSYPFF 900
QY 901 RAKAKVAPFLDSEFSKIFLHLEIELEAAGSDSNERDSTKEDNVAPLRFHLKYEADVLFT 960
Db 901 RAKAKVAPFLDSEFSKIFLHLEIELEAAGSDSNERDSTKEDNVAPLRFHLKYEADVLFT 960
QY 961 RSSSLSHYEVKUNSLERYDGTGPPFCIFRQNLGLPIHGMKMITIPIATRSGNRLL 1020
Db 961 RSSSLSHYEVKUNSLERYDGTGPPFCIFRQNLGLPIHGMKMITIPIATRSGNRLL 1020
QY 1021 KURDFLTDEANTSCNIWGNSTERYPTPVEEDLRRAPOLNHSNDVVSINCLRLVNPQEI 1080
Db 1021 KURDFLTDEANTSCNIWGNSTERYPTPVEEDLRRAPOLNHSNDVVSINCLRLVNPQEI 1080
QY 1081 NPHLLGNLWLSLKALKYKSMKIMVNAALQROFHSPPIFREEDPSRQIEFEISKQEDWQV 1140
Db 1081 NPHLLGNLWLSLKALKYKSMKIMVNAALQROFHSPPIFREEDPSRQIEFEISKQEDWQV 1140
QY 1141 PIWIIVGSTLGLLLALLVIALKLGFPSARRRRRPGLDPTPKVLE 1188
Db 1141 PIWIIVGSTLGLLLALLVIALKLGFPSARRRRRPGLDPTPKVLE 1188
```

```
RESULT 2
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: NO. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-810

Query Match 99.5%; Score 6192; DB 15; Length 1188;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1183; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MDLPRLVAVALSILWPGFTDTFNMDTRKPRVIPSRTAFPGTYTVOOHDISGNKWLAVGA 60
Db 1 MDLPRLVAVALSILWPGFTDTFNMDTRKPRVIPSRTAFPGTYTVOOHDISGNKWLAVGA 60
QY 61 PLETTYQKTDGVYKCPVIHGNCTKLNLRVTLSNVSRKKNMRGLSLATNPKDNSFLA 120
Db 61 PLETTYQKTDGVYKCPVIHGNCTKLNLRVTLSNVSRKKNMRGLSLATNPKDNSFLA 120
QY 121 CSPLWSHECGSSYYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSNYSYPWVE 180
Db 121 CSPLWSHECGSSYYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSNYSYPWVE 180
QY 181 VOHFLINILKKFYIGPGQIQGVVQYGEDVHVEHFLNDYRSVKVDVVEAASHIEQGGTET 240
Db 181 VOHFLINILKKFYIGPGQIQGVVQYGEDVHVEHFLNDYRSVKVDVVEAASHIEQGGTET 240
QY 241 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHSDSPDLEKVIQOQSERDNVTRYAVAVL 300
Db 241 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHSDSPDLEKVIQOQSERDNVTRYAVAVL 300
QY 301 GYNNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDIDVALGDRIFSLEGTNKNET 360
Db 301 GYNNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDIDVALGDRIFSLEGTNKNET 360
QY 361 SFGLEMSOTGFSSSHVVVEDGVLGAVGYDNGAVILKETSAKVIPLESYLYKEFPEELKN 420
Db 361 SFGLEMSOTGFSSSHVVVEDGVLGAVGYDNGAVILKETSAKVIPLESYLYKEFPEELKN 420
QY 421 HGAYLGYTVTSVVSRRQGRVYVAGAPRNFHTGKVLFTMHNRSITIHQAMRGOQIGSYF 480
Db 421 HGAYLGYTVTSVVSRRQGRVYVAGAPRNFHTGKVLFTMHNRSITIHQAMRGOQIGSYF 480
QY 481 GSEITSVSDIDGCVTDVLLVGAPMYFNEGRGKVVYVELRQNRVYNGTLKDSHYQNA 540
Db 481 GSEITSVSDIDGCVTDVLLVGAPMYFNEGRGKVVYVELRQNRVYNGTLKDSHYQNA 540
QY 541 RFGSSIASVRDLNQSDYNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQRTITASELATG 600
Db 541 RFGSSIASVRDLNQSDYNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQRTITASELATG 600
QY 601 LQYFGCSIHGQJDLNEDGLIDLAVGALGNVILWSRPVVQINASHLHFEPKINIFHRDCK 660
```

Db 601 LQYFGCSIHGQDLNEDGLIDAVGALGNVILWSRPVQNASLHFFSPKINIFHRDCK 660
QY 661 RSGRDATCLAAFLCFTPIFLAPHFQTTVGIRYNATMDERRYTPRAHLDGEGDFTNRVAV 720
Db 661 RSGRDATCLAAFLCFTPIFLAPHFQTTVGIRYNATMDERRYTPRAHLDGEGDFTNRVAV 720
QY 721 LSSGQELCERINFHVLDTDYVVPVTSVEYSLEDPDHGPMDDGWPTTLRVSPVFWNG 780
Db 721 LSSGQELCERINFHVLDTDYVVPVTSVEYSLEDPDHGPMDDGWPTTLRVSPVFWNG 780
QY 781 CNEDEHCVPLDVLDAARSDLPAMEYCORVLRKPAQDCSAYTSLSDTTVFIIESTRQAV 840
Db 781 CNEDEHCVPLDVLDAARSDLPAMEYCORVLRKPAQDCSAYTSLSDTTVFIIESTRQAV 840
QY 841 EATLENRGENAYSTVLNISOSANLOFASLIQKEDSDGSEICVNEERRLQKQVNVSPYFF 900
Db 841 EATLENRGENAYSTVLNISOSANLOFASLIQKEDSDGSEICVNEERRLQKQVNVSPYFF 900
QY 901 RAKAKVAFRLDSEFSKSIPLHLEIEIAAGSDSNBERDSTKEDNVAPLRFHLKYEADVLT 960
Db 901 RAKAKVAFRLDSEFSKSIPLHLEIEIAAGSDSNBERDSTKEDNVAPLRFHLKYEADVLT 960
QY 961 RSSLSHYEVKLNSSLERYDGIQPPFSCIPRIQNLGLFPIHGMMKTIPIATRSNRL 1020
Db 961 RSSLSHYEVKLNSSLERYDGIQPPFSCIPRIQNLGLFPIHGMMKTIPIATRSNRL 1020
QY 1021 KLRFDLTDEANTSCNIWGNSTYRPTVPEEDLRAPQNLHNSDVSINCNIRLVPQEI 1080
Db 1021 KLRFDLTDEANTSCNIWGNSTYRPTVPEEDLRAPQNLHNSDVSINCNIRLVPQEI 1080
QY 1081 NFHLLGNLWLSLKALKYKSMKIWNALQRFHSPIFREEDPSRQIEFISKOEDQV 1140
Db 1081 NFHLLGNLWLSLKALKYKSMKIWNALQRFHSPIFREEDPSRQIEFISKOEDQV 1140
QY 1141 PIWIVGSTGLLLALLVLALRKLGFRRSARRRREGLDPTPKVLE 1188
Db 1141 PIWIVGSTGLLLALLVLALRKLGFRRSARRRREGLDPTPKVLE 1188

RESULT 3

US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984.130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-35

Query Match 99.4%; Score 6188.5; DB 10; Length 1189;
Best Local Similarity 99.6%; Pred. No. 0;

Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDLPRGLVVAWALSIPGFTDTFNMDRKPRVIFGSRRTAFGYTVQQHDSIGNKMLVGA 60

Db 1 MDLPRGLVVAWALSIPGFTDTFNMDRKPRVIFGSRRTAFGYTVQQHDSIGNKMLVGA 60
QY 61 PLENGYQKTDVYKCPVHGNCTKLNLRGRTLSNVSEKDNMRGLSLATNPKNDSFLA 120
Db 61 PLENGYQKTDVYKCPVHGNCTKLNLRGRTLSNVSEKDNMRGLSLATNPKNDSFLA 120
QY 121 CSPLWSHCSSYYTGTMCSSRVNSNFRFSKTVAALQRCQTYMDIIVILVDSGNSIYPWVE 180
Db 121 CSPLWSHCSSYYTGTMCSSRVNSNFRFSKTVAALQRCQTYMDIIVILVDSGNSIYPWVE 180
QY 181 VOHFLINILKFFYIGPGQIQGVVQYGEDVHVEHLDNDRSVKDVWEAASHIEORGGET 240
Db 181 VOHFLINILKFFYIGPGQIQGVVQYGEDVHVEHLDNDRSVKDVWEAASHIEORGGET 240
QY 241 RTAFGIETSEAFQKGRKGAKKVMIVITDGEHSDSPDLEKVIQOQSERDNTVYAVAVL 300
Db 241 RTAFGIETSEAFQKGRKGAKKVMIVITDGEHSDSPDLEKVIQOQSERDNTVYAVAVL 300
QY 301 GYNNRGINPBTFFNEIKYIATSDPDHGFNVNTDEAALKDVIDALGDRIFPLEGNTKNET 360
Db 301 GYNNRGINPBTFFNEIKYIATSDPDHGFNVNTDEAALKDVIDALGDRIFPLEGNTKNET 360
QY 361 SFGLEMSOTGFSHVVEDGVLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEPPELKN 420
Db 361 SFGLEMSOTGFSHVVEDGVLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEPPELKN 420
QY 421 HGAYLGYTVTISVSSRQGRVVVAGAPRNFHTGKVFLETMHNNRSLTIHQAMRGQOIGSYF 480
Db 421 HGAYLGYTVTISVSSRQGRVVVAGAPRNFHTGKVFLETMHNNRSLTIHQAMRGQOIGSYF 480
QY 481 GSEITSDVIDGDGVTDLVLCAPMYNEGERBGRKVYVELRQNPVFNVTGTLKDSHSYQNA 540
Db 481 GSEITSDVIDGDGVTDLVLCAPMYNEGERBGRKVYVELRQNPVFNVTGTLKDSHSYQNA 540
QY 541 RFGSSIASVRLNODSYNDVVVGPAPLSDNHAGAIYIFHGPRGSLTKTPKQITASELATG 600
Db 541 RFGSSIASVRLNODSYNDVVVGPAPLSDNHAGAIYIFHGPRGSLTKTPKQITASELATG 600
QY 601 LQYFGCSIHGQDLNEDGLIDAVGALGNVILWSRPVQNASLHFFSPKINIFHRDCK 660
Db 601 LQYFGCSIHGQDLNEDGLIDAVGALGNVILWSRPVQNASLHFFSPKINIFHRDCK 660
QY 661 RSGRDATCLAAFLCFTPIFLAPHFQTTVGIRYNATMDERRYTPRAHLDGEGDFTNRVAV 720
Db 661 RSGRDATCLAAFLCFTPIFLAPHFQTTVGIRYNATMDERRYTPRAHLDGEGDFTNRVAV 720
QY 721 LSSGQELCERINFHVLDTDYVVPVTSVEYSLEDPDHGPMDDGWPTTLRVSPVFWNG 780
Db 721 LSSGQELCERINFHVLDTDYVVPVTSVEYSLEDPDHGPMDDGWPTTLRVSPVFWNG 780
QY 781 CNEDEHCVPLDVLDAARSDLPAMEYCORVLRKPAQDCSAYTSLSDTTVFIIESTRQAV 840
Db 781 CNEDEHCVPLDVLDAARSDLPAMEYCORVLRKPAQDCSAYTSLSDTTVFIIESTRQAV 840
QY 841 EATLENRGENAYSTVLNISOSANLOFASLIQKEDSDGSEICVNEERRLQKQVNVSPYFF 900
Db 841 EATLENRGENAYSTVLNISOSANLOFASLIQKEDSDGSEICVNEERRLQKQVNVSPYFF 900
QY 901 RAKAKVAFRLDSEFSKSIPLHLEIEIAAGSDSNBERDSTKEDNVAPLRFHLKYEADVLT 960
Db 901 RAKAKVAFRLDSEFSKSIPLHLEIEIAAGSDSNBERDSTKEDNVAPLRFHLKYEADVLT 960
QY 961 RSSLSHYEVKLNSSLERYDGIQPPFSCIPRIQNLGLFPIHGMMKTIPIATRSNRL 1020
Db 961 RSSLSHYEVKLNSSLERYDGIQPPFSCIPRIQNLGLFPIHGMMKTIPIATRSNRL 1020
QY 1021 KLRFDLTDEANTSCNIWGNSTYRPTVPEEDLRAPQNLHNSDVSINCNIRLVPQEI 1079
Db 1021 KLRFDLTDEANTSCNIWGNSTYRPTVPEEDLRAPQNLHNSDVSINCNIRLVPQEI 1080
QY 1080 INFHLLGNLWLSLKALKYKSMKIWNALQRFHSPIFREEDPSRQIEFISKOEDQV 1139

Db 1081 INFHLLGNLWLSRKALKYKSMKIMVNAALQRFHSPFIREDPSRQIVFEISKQEDWQ 1140

QY 1140 VPIWIIIVGSTGLGALLLALLVIALKLGKGFRRSARRRPPGLDPTPKVLE 1188

Db 1141 VPIWIIIVGSTGLGALLLALLVIALWKLGFRRSARRRPPGLDPTPKVLE 1189

RESULT 4

US-09-836-353A-35

; Sequence 35, Application US/09836353A

; Publication No. US20030129685A1

; GENERAL INFORMATION:

; APPLICANT: Ni et al.

; TITLE OF INVENTION: 12 Human Secreted Proteins

; FILE REFERENCE: PF489PI

; CURRENT APPLICATION NUMBER: US/09/836,353A

; CURRENT FILING DATE: 2001-04-18

; PRIOR APPLICATION NUMBER: 60/198,407

; PRIOR FILING DATE: 2000-04-19

; PRIOR APPLICATION NUMBER: PCT/US99/25031

; PRIOR FILING DATE: 1999-10-27

; PRIOR APPLICATION NUMBER: 60/105,971

; PRIOR FILING DATE: 1998-10-28

; NUMBER OF SEQ ID NOS: 147

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 35

; LENGTH: 1189

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-836-353A-35

Query Match 99.4%; Score 6188.5; DB 10; Length 1189;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MDLPRGLVVAWALSMPGFTDFTNMTRKPRVPGSRVTAFFGYTQQHDSGNKWLVGGA 60

Db 1 MDLPRGLVVAWALSMPGFTDFTNMTRKPRVPGSRVTAFFGYTQQHDSGNKWLVGGA 60

QY 61 PLETNGYQKTDGVYKCPVHIGNCTKLNLRVTLSNVSEKDNRLGLSLATNPKNSFLA 120

Db 61 PLETNGYQKTDGVYKCPVHIGNCTKLNLRVTLSNVSEKDNRLGLSLATNPKNSFLA 120

QY 121 CSPLSHSCGSSVYTTGMCSSRVNSNPRFSKTVPALQRCQTYMDIVIVLDGNSYYPWE 180

Db 121 CSPLSHSCGSSVYTTGMCSSRVNSNPRFSKTVPALQRCQTYMDIVIVLDGNSYYPWE 180

QY 181 VQHFLNLLKXYIGPGQIQGVVQYGEDVWHEFHNDYRSKDVVVEAAASHIEQRGGTET 240

Db 181 VQHFLNLLKXYIGPGQIQGVVQYGEDVWHEFHNDYRSKDVVVEAAASHIEQRGGTET 240

QY 241 RTAFGIEFARSAFQGRKGAKKVMIVITDGSHDSPLEKVIQOSERDNTVRVAVL 300

Db 241 RTAFGIEFARSAFQGRKGAKKVMIVITDGSHDSPLEKVIQOSERDNTVRVAVL 300

QY 301 GYNNRGINPETFLEIKVIASDPDKHFNVTDEAALXDIIDALGDRIFSLGNTKNET 360

Db 301 GYNNRGINPETFLEIKVIASDPDKHFNVTDEAALXDIIDALGDRIFSLGNTKNET 360

QY 361 SFGLEMSQTFSSHVVEDGVLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEFPEELKN 420

Db 361 SFGLEMSQTFSSHVVEDGVLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEFPEELKN 420

QY 421 HGAYLGTYTTSVSSRQGRVYVAGAPRPHNTGKVILFTWHNRSRLTIHOAMGQQIGSYF 480

Db 421 HGAYLGTYTTSVSSRQGRVYVAGAPRPHNTGKVILFTWHNRSRLTIHOAMGQQIGSYF 480

QY 481 GSEITSVDIDGCVTDVLLVGAPMYNEGRGKVVVYELQNRVYNGTLKDSHSYQNA 540

Db 481 GSEITSVDIDGCVTDVLLVGAPMYNEGRGKVVVYELQNRVYNGTLKDSHSYQNA 540

QY 541 RFGSSIASVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGRGSIILKTPKQRTITASELATG 600

RESULT 5

US-10-262-839-4

; Sequence 4, Application US/10262839

; Publication No. US20040038877A1

; GENERAL INFORMATION:

; APPLICANT: Alsobrook, John,

; APPLICANT: Anderson, David W.,

; APPLICANT: Boldog, Ferenc,

; APPLICANT: Burgesse, Catherine,

; APPLICANT: Catterton, Elina,

; APPLICANT: Edinger, Shlomit,

; APPLICANT: Ellerman, Karen,

; APPLICANT: Gerlach, Valerie,

; APPLICANT: Gorman, Linda,

; APPLICANT: Guo, Xiaojia,

; APPLICANT: Ji, Weizhen,

; APPLICANT: Kekuda, Ramesh,

; APPLICANT: Leach, Martin,

; APPLICANT: Li, Li,

; APPLICANT: Miller, Charles,

; APPLICANT: Patturajan, Meera,

; APPLICANT: Reiger, Daniel,

; APPLICANT: Rothenberg, Mark,

; APPLICANT: Shinkets, Richard,

; APPLICANT: Smithson, Glennnda,

; APPLICANT: Spytek, Kimberly,

; APPLICANT: Taupier, Raymond, jr.,

; APPLICANT: Vernet, Corine,

; APPLICANT: Voss, Edward,

; APPLICANT: Zerhusen, Brian,

APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-4

Query March 99.4%; Score 6188.5; DB 12; Length 1189;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1184; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

1 MDLPRGLVVAWALSLWPGFTDTEHMDTRKPRVPGSRTAFPGYTVQOHDISGNKWLTVGA 60
1 MDLPRGLVVAWALSLWPGFTDTEHMDTRKPRVPGSRTAFPGYTVQOHDISGNKWLTVGA 60
61 PLEINGYQKTGDVYKCPVGHGNCIKNLGRVTLNSVSRKDNMRLGLSLATNPKNDSFLA 120
61 PLEINGYQKTGDVYKCPVGHGNCIKNLGRVTLNSVSRKDNMRLGLSLATNPKNDSFLA 120
121 CSPLWSHECGSSYTTGMCGRVNSNFRPSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
121 CSPLWSHECGSSYTTGMCGRVNSNFRPSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
181 VQHEFLINILKFKYIGPQIQGVVQYGEDVVEHFLNDYRSVKDVVEAASHIEQGGTET 240
181 VQHEFLINILKFKYIGPQIQGVVQYGEDVVEHFLNDYRSVKDVVEAASHIEQGGTET 240
241 RTAFGIEFARSEAFQKGRGKAKVMIVITDGHSDSPDLEKVIQOGERDNVTVYAVVL 300
241 RTAFGIEFARSEAFQKGRGKAKVMIVITDGHSDSPDLEKVIQOGERDNVTVYAVVL 300
301 GYNNRGINETFLNEIKYASDPDDKGFHFNVTDEAALKOIVDALGRIFSLGNTKNET 360
301 GYNNRGINETFLNEIKYASDPDDKGFHFNVTDEAALKOIVDALGRIFSLGNTKNET 360
361 SFGLMSQTFSSHVWEDGVLGAVGAYDNAGVLTSAKVIPLRESYLKPEPELKN 420
361 SFGLMSQTFSSHVWEDGVLGAVGAYDNAGVLTSAKVIPLRESYLKPEPELKN 420
421 HGAYLGYTVTSVSSRGRVYVAGAPRNFHTGKVIPLTMNNRSLTIHQMRGQIGSYF 480
421 HGAYLGYTVTSVSSRGRVYVAGAPRNFHTGKVIPLTMNNRSLTIHQMRGQIGSYF 480
481 GSEITSVIDDGDGVTDLVLLGAPYFNEGRGKVVYELRQNFVYNGTLKDSHSYQNA 540
481 GSEITSVIDDGDGVTDLVLLGAPYFNEGRGKVVYELRQNFVYNGTLKDSHSYQNA 540

QY 541 RFGSSIASVRDLNQSYNDVVVVGAPLENDHAGAIYIFHGFRGSILTKPKQITASELATG 600
Db 541 RFGSSIASVRDLNQSYNDVVVVGAPLENDHAGAIYIFHGFRGSILTKPKQITASELATG 600
QY 601 LOYFGCSIHGOLDNEDGLIDLAVGALGNVILSRPVVQINASLHFFPSKINIFHRDCK 660
Db 601 LOYFGCSIHGOLDNEDGLIDLAVGALGNVILSRPVVQINASLHFFPSKINIFHRDCK 660
QY 661 RSGRDATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYTPRAHLDGEGDFTNRVAV 720
Db 661 RSGRDATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYTPRAHLDGEGDFTNRVAV 720
QY 721 LLSGQELCERINHFVLDTDADYVVPVTFSEYSLSDPDHGMPLDGMPTTLRVSVPFWNG 780
Db 721 LLSGQELCERINHFVLDTDADYVVPVTFSEYSLSDPDHGMPLDGMPTTLRVSVPFWNG 780
QY 781 CNEDEHCVDPDLVDARSDLPAMEYCORVLRKPQDCSAYTLSPTDTTVFIESTRQVRVAV 840
Db 781 CNEDEHCVDPDLVDARSDLPAMEYCORVLRKPQDCSAYTLSPTDTTVFIESTRQVRVAV 840
QY 841 EATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNEERLQKQVCNVSYPPF 900
Db 841 EATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNEERLQKQVCNVSYPPF 900
QY 901 RAKAKVAFRLDSEFSKSIIFLHLHLEIILAAGSDSNERDSTKEDNVAPLRFLHLYEADVLFT 960
Db 901 RAKAKVAFRLDSEFSKSIIFLHLHLEIILAAGSDSNERDSTKEDNVAPLRFLHLYEADVLFT 960
QY 961 RSSLSHYEVKLNSSLERYDGIGPPFSCIFRIQNLGLFPIHGMNMKITIPIATSGNELL 1020
Db 961 RSSLSHYEVKLNSSLERYDGIGPPFSCIFRIQNLGLFPIHGMNMKITIPIATSGNELL 1020
QY 1021 KLDRDLTDE-ANTSCTIWNSTYRPTFVEEDLERAPQLAHNSDVSINCNIRLVPNOE 1079
Db 1021 KLDRDLTDEANTSCTIWNSTYRPTFVEEDLERAPQLAHNSDVSINCNIRLVPNOE 1080
QY 1080 INFHLLGNLWLSLKALKYKSMKIMVNAALQROPHSPFIIFREDDPSQIBFEISKQEDWQ 1139
Db 1081 INFHLLGNLWLSLKALKYKSMKIMVNAALQROPHSPFIIFREDDPSQIBFEISKQEDWQ 1140
QY 1140 VPTWIIWGSTLGGLLLLALLVLRKLGFRRSRRRRREPGLDTPPKYLE 1188
Db 1141 VPTWIIWGSTLGGLLLLALLVLRKLGFRRSRRRRREPGLDTPPKYLE 1189

RESULT 6
US-10-262-839-6
; Sequence 6, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Smithson, Richard,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spyttek, Kimberly,
; APPLICANT: Taupier, Raymond, jr.,

APPLICANT: Vernet, Corine,
APPLICANT: Voss, Edward,
APPLICANT: Zernusen, Brian,
APPLICANT: Zhong, Mei
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
FILE REFERENCE: 21402-462A
CURRENT APPLICATION NUMBER: US/10/262,839
CURRENT FILING DATE: 2002-10-01
PRIOR APPLICATION NUMBER: 60/326,483
PRIOR FILING DATE: 2001-10-02
PRIOR APPLICATION NUMBER: 60/327,917
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/328,029
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/328,056
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/381,101
PRIOR FILING DATE: 2002-05-16
PRIOR APPLICATION NUMBER: 60/371,972
PRIOR FILING DATE: 2002-04-12
PRIOR APPLICATION NUMBER: 60/327,342
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 60/328,044
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/328,849
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/374,738
PRIOR FILING DATE: 2002-04-23
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 367
SOFTWARE: CuraSeqList version 0.1
SEQ ID NO 6
LENGTH: 1120
TYPE: PRT
ORGANISM: Homo sapiens
US-10-262-839-6

Query Match 92.9%; Score 5780; DB 12; Length 1120;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1115; Conservative 1; Mismatches 3; Indels 70; Gaps 2;

QY 1 MDLPRGLVAVALSMPGFTDTFNMDTRKPRVPGSRPTAFQYTVQQHDISGNKWLAVGA 60
Db 1 MDLPRGLVAVALSMPGFTDTFNMDTRKPRVPGSRPTAFQYTVQQHDISGNKWLAVGA 60
QY 61 PLETNGYQKTGVYKCPVIHGNCTKLNIGRVTLNVNRSERKDNRLGLSLATNPKNSFLA 120
Db 61 PLETNGYQKTGVYKCPVIHGNCTKLNIG----- 89
QY 121 CSPLNSHEGSSYYITGMCSRNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVE 180
Db 90 -----CQTYMDIVIVLDGNSIYPWVE 111
QY 181 VOHFLINILKFYIGPGQIQGVQVQYGEDVWHEFHNDYRSKDVVEAASHIEQRGGTET 240
Db 112 VOHFLINILKFYIGPGQIQGVQVQYGEDVWHEFHNDYRSKDVVEAASHIEQRGGTET 171
QY 241 RTAFGIEFARSFAQGGKRGAKKUMIVITDGESHSDPLEKVIQOESRDNTRYAVAVL 300
Db 172 RTAFGIEFARSFAQGGKRGAKKUMIVITDGESHSDPLEKVIQOESRDNTRYAVAVL 231
QY 301 GYNNRGINPETFLNEIKVIASDPDDKHFNFVNTDEAALKDIVDALGDRIFSLEGTKNET 360
Db 232 GYNNRGINPETFLNEIKVIASDPDDKHFNFVNTDEAALKDIVDALGDRIFSLEGTKNET 291
QY 361 SFGLEMSQTFGSHVVEDGVILGAVYDNGAVLKETSAGKVIPLRESYLKEFPPELKN 420
Db 292 SFGLEMSQTFGSHVVEDGVILGAVYDNGAVLKETSAGKVIPLRESYLKEFPPELKN 351
QY 421 HGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTWHNRSRLTIHOAMRGQQIGSYF 480
Db 352 HGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTWHNRSRLTIHOAMRGQQIGSYF 411

QY 481 GSEITSDIDGDTVDLLVVGAPMYFNEGRGRGVVYVELRQNFVYNGTLKDSHSYQNA 540
Db 412 GSEITSDIDGDTVDLLVVGAPMYFNEGRGRGVVYVELRQNFVYNGTLKDSHSYQNA 471
QY 541 RFGSSIASVRDLNDSYNDVVVVGAPLEDNHAAGIYIFHGFRGSILKTPKQITASELATG 600
Db 472 RFGSSIASVRDLNDSYNDVVVVGAPLEDNHAAGIYIFHGFRGSILKTPKQITASELATG 531
QY 601 LOYFGCSIHGQDLINEDGLIDLAVGALGNVILSRPVQVQINASHFEPSPKINIPHRDCK 660
Db 532 LOYFGCSIHGQDLINEDGLIDLAVGALGNVILSRPVQVQINASHFEPSPKINIPHRDCK 591
QY 661 RSGRDATCLAAFLCFTPIFLAPHFQTTVIGIRYNATMDERRYTPRAHLDGGRFTNRAV 720
Db 592 RSGRDATCLAAFLCFTPIFLAPHFQTTVIGIRYNATMDERRYTPRAHLDGGRFTNRAV 651
QY 721 LLSGQELCERINFHVLDTADYVVPVTFPSVEYSLEDPHGMLDDGWPVTLRVSVVPFWNG 780
Db 652 LLSGQELCERINFHVLDTADYVVPVTFPSVEYSLEDPHGMLDDGWPVTLRVSVVPFWNG 711
QY 781 CNEDEHCVPLDVLDAARSDLPAMEYQCVLRKPAQDCSAYTILSPDTTTFIIESTRQVAV 840
Db 712 CNEDEHCVPLDVLDAARSDLPAMEYQCVLRKPAQDCSAYTILSPDTTTFIIESTRQVAV 771
QY 841 EATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNBERRLQKQVNSYPPFF 900
Db 772 EATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNBERRLQKQVNSYPPFF 831
QY 901 RAKAKVAPRLDSEFSKSIHLHLEIELAAGSDSNERDSTKEDNVAPLFLHLYEADVLFT 960
Db 832 RAKAKVAPRLDSEFSKSIHLHLEIELAAGSDSNERDSTKEDNVAPLFLHLYEADVLFT 891
QY 961 RSSLSHVEVKLNSLERYDGIQPPFSCIFRIONLGLPFIHGMKMIITPIATRSGNRL 1020
Db 892 RSSLSHVEVKLNSLERYDGIQPPFSCIFRIONLGLPFIHGMKMIITPIATRSGNRL 951
QY 1021 KLRDLTDE-ANTSNIWGNSTEXRPTVEEDLRAPQLNHSNDSVVSINCLRLVPNQE 1079
Db 952 KLRDLTDEVANTSCNIWGNSTEXRPTVEEDLRAPQLNHSNDSVVSINCLRLVPNQE 1011
QY 1080 INFHLLGNLWLRSLKALKYKSMKIMVNAALORQHSPIFREEDSRQIEFISKQEDWQ 1139
Db 1012 INFHLLGNLWLRSLKALKYKSMKIMVNAALORQHSPIFREEDSRQIEFISKQEDWQ 1071
QY 1140 VPIIIVGSTIGLLLLALLVLLALRKLGFERSARRRRRPGDPTPKVLE 1188
Db 1072 VPIIIVGSTIGLLLLALLVLLALRKLGFERSARRRRRPGDPTPKVLE 1120

RESULT 7
US-09-984-130-43
Sequence 43, Application US/09984130
Publication No. US2003005231A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489P2
CURRENT APPLICATION NUMBER: US/09/984,130
CURRENT FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 60/243,792
PRIOR FILING DATE: 2000-10-30
PRIOR APPLICATION NUMBER: 09/836,353
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 149
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 43
LENGTH: 1034

Db 541 RFGSSIASVRDLNQSYNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQITASELATG 600
QY 601 LQYFGCSIHGOLDNEDGLIDLAVGALGNVILWSRPVVQINASHLHFPSPKINIPHRDCK 660
Db 601 LQYFGCSIHGOLDNEDGLIDLAVGALGNVILWSRPVVQINASHLHFPSPKINIPHRDCK 660
QY 661 RSGRDATCLAAFLCPTPIFLAPHFQTTTGGIRYNATMDERRYTTPRAHLDEGDRFTNRV 720
Db 661 RSGRDATCLAAFLCPTPIFLAPHFQTTTGGIRYNATMDERRYTTPRAHLDEGDRFTNRV 720
QY 721 LLSGQELCERINHVLTADVVKPVSVEVSLDDPHGMLDGMPTTLRVSPFWNG 780
Db 721 LLSGQELCERINHVLTADVVKPVSVEVSLDDPHGMLDGMPTTLRVSPFWNG 780
QY 781 CNEDEHCVFDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTLSTFTTFFIESTRQRAV 840
Db 781 CNEDEHCVFDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTLSTFTTFFIESTRQRAV 840
QY 841 EATLENGENAYSTVNLISQSANLOFASLIQKEDSDGIECVNEERLQKQVCNVSYPFF 900
Db 841 EATLENGENAYSTVNLISQSANLOFASLIQKEDSDGIECVNEERLQKQVCNVSYPFF 900
QY 901 RAKAVAFELDSERFSKIFLHLELELAAGSDSNERDSTKEDNVAPLRFHLKYEADVLT 960
Db 901 RAKAVAFELDSERFSKIFLHLELELAAGSDSNERDSTKEDNVAPLRFHLKYEADVLT 960
QY 961 RSSLSHYEVKLNSSLRYDGGIPPFSCIFRIONLGLPFIHGMKMITIPIATRSNRL 1020
Db 961 RSSLSHYEVKLNSSLRYDGGIPPFSCIFRIONLGLPFIHGMKMITIPIATRSNRL 1020
QY 1021 KLRDLTDE 1029
Db 1021 KLRDLTDE 1029

RESULT 9

US-09-764-870-313
; Sequence 313, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P7214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-313

Query Match 58.2%; Score 3621; DB 9; Length 707;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 689; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLPRLGVAVWALSLLWPGFTDTFNMDTRKPRVPGSRTPAFGTYVQQHDSIGNKWLWGA 60
Db 17 MDLPRLGVAVWALSLLWPGFTDTFNMDTRKPRVPGSRTPAFGTYVQQHDSIGNKWLWGA 76
QY 61 PLETNGYQKTGDVYKCPVHGNCTKLNLRVTLNSVSRKDNMRGLSLATNPKNSFLA 120
Db 77 PLETNGYQKTGDVYKCPVHGNCTKLNLRVTLNSVSRKDNMRGLSLATNPKNSFLA 136
QY 121 CSPLWSHECGSSYYTTGMCSSRVNSNFRSKTVAPALQRCQTYMDIVILVDGNSIYPWVE 180
Db 137 CSPLWSHECGSSYYTTGMCSSRVNSNFRSKTVAPALQRCQTYMDIVILVDGNSIYPWVE 196
QY 181 VOHFLINILKKFYIGPGQIQGVVQYGEDVHFEHLNDYRSVKDVVEAAASHIEQRGGTET 240

Db 197 VOHFLINILKKFYIGPGQIQGVVQYGEDVHFEHLNDYRSVKDVVEAAASHIEQRGGTET 256
QY 241 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGEHSDSPLEKVIQQSRDNNTRYAVAVL 300
Db 257 RTAFGIEFARSEAFQKGRKGAKKVMIVITDGEHSDSPLEKVIQQSRDNNTRYAVAVL 316
QY 301 GYNNRRGINPETFLNEIKYIASDDDDKHFFNVDTDEAALKDIDVALGDRIFSLGEGNKNET 360
Db 317 GYNNRRGINPETFLNEIKYIASDDDDKHFFNVDTDEAALKDIDVALGDRIFSLGEGNKNET 376
QY 361 SFGLEMSQTGFSSHVVEDGVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEPPEELKN 420
Db 377 SFGLEMSQTGFSSHVVEDGVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEPPEELKN 436
QY 421 HGAYLGYTIVTSVSSRQGRVVVAGAPRNHTGKULLFTMHNNRSITIHQAMRGQOIGSYF 480
Db 437 HGAYLGYTIVTSVSSRQGRVVVAGAPRNHTGKULLFTMHNNRSITIHQAMRGQOIGSYF 496
QY 481 GSEITSDVIDDGDGVTDVLLVGAPMYFNEGRGRGVVYVELRQNRVYVNGTLKDSHYQNA 540
Db 497 GSEITSDVIDDGDGVTDVLLVGAPMYFNEGRGRGVVYVELRQNLFPVYNGTLKDSHYQNA 556
QY 541 RFGSSIASVRDLNQSYNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQITASELATG 600
Db 557 RFGSSIASVRDLNQSYNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQITASELATG 616
QY 601 LQYFGCSIHGOLDNEDGLIDLAVGALGNVILWSRPVVQINASHLHFPSPKINIPHRDCK 660
Db 617 LQYFGCSIHGOLDNEDGLIDLAVGALGNVILWSRPVVQINASHLHFPSPKINIPHRDCK 676
QY 661 RSGRDATCLAAFLCPTPIFLAPHFQTTTGI 691
Db 677 RSGRDATCLAAFLCPTPIFLAPHFQTTTGI 707

RESULT 10

US-10-125-540-313
; Sequence 313, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214C1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-540-313

Query Match 58.2%; Score 3621; DB 14; Length 707;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 689; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLPRLGVAVWALSLLWPGFTDTFNMDTRKPRVPGSRTPAFGTYVQQHDSIGNKWLWGA 60
Db 17 MDLPRLGVAVWALSLLWPGFTDTFNMDTRKPRVPGSRTPAFGTYVQQHDSIGNKWLWGA 76
QY 61 PLETNGYQKTGDVYKCPVHGNCTKLNLRVTLNSVSRKDNMRGLSLATNPKNSFLA 120
Db 77 PLETNGYQKTGDVYKCPVHGNCTKLNLRVTLNSVSRKDNMRGLSLATNPKNSFLA 136
QY 121 CSPLWSHECGSSYYTTGMCSSRVNSNFRSKTVAPALQRCQTYMDIVILVDGNSIYPWVE 180
Db 137 CSPLWSHECGSSYYTTGMCSSRVNSNFRSKTVAPALQRCQTYMDIVILVDGNSIYPWVE 196
QY 181 VOHFLINILKKFYIGPGQIQGVVQYGEDVHFEHLNDYRSVKDVVEAAASHIEQRGGTET 240
Db 197 VOHFLINILKKFYIGPGQIQGVVQYGEDVHFEHLNDYRSVKDVVEAAASHIEQRGGTET 256

QY 241 RTAFGIEPARSAFOKGRGKAKKVMIVITDGHSDSPDLKVIQOESRDNTRYAVAVL 300
DB 257 RTAFGIEPARSAFOKGRGKAKKVMIVITDGHSDSPDLKVIQOESRDNTRYAVAVL 316
QY 301 GYNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAKDIIVDALGDRIFSLEGTNKNET 360
DB 317 GYNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAKDIIVDALGDRIFSLEGTNKNET 376
QY 361 SFGLEMSQTFSSHVEDGVILGAVGAYDNGAVILKETSAGKVIPIRESYLKEFFPEELKN 420
DB 377 SFGLEMSQTFSSHVEDGVILGAVGAYDNGAVILKETSAGKVIPIRESYLKEFFPEELKN 436
QY 421 HGAYLGYTVTSVSSRQGRVVVAGAPRFNHGTGVILFTWHNRSLTIHQAMSGQOIGSYF 480
DB 437 HGAYLGYTVTSVSSRQGRVVVAGAPRFNHGTGVILFTWHNRSLTIHQAMSGQOIGSYF 496
QY 481 GSEITSDVDIGDVTDLVAVGAPMYFNEGRERGVVYVELRQNFVYNGTLDKSHSYQNA 540
DB 497 GSEITSDVDIGDVTDLVAVGAPMYFNEGRERGVVYVELRQNFVYNGTLDKSHSYQNA 556
QY 541 REGSSIASVRDLNQSYNDVVVGAPELDNHAGAIYIFHGFGRSILKTPKORITASELATG 600
DB 557 REGSSIASVRDLNQSYNDVVVGAPELDNHAGAIYIFHGFGRSILKTPKORITASELATG 616
QY 601 LQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSPVQVQINASHLHFEPSKINIFHRDCK 660
DB 617 LQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSPVQVQINASHLHFEPSKINIFHRDCK 676
QY 661 RSRGDRATCLAAFLCFTPIFLAPHFQTTVGI 691
DB 677 RSRGDRATCLAAFLCFTPIFLAPHFQTTVGI 707

RESULT 11

US-09-866-050A-624
; Sequence 624, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:

; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-624

Query Match 51.1%; Score 3182; DB 10; Length 688;
Best Local Similarity 86.5%; Pred. No. 3.1e-284;
Matches 595; Conservative 46; Mismatches 47; Indels 0; Gaps 0;

QY 501 GAPMYFNEGRERGVVYVELRQNFVYNGTLDKSHSYQNAFPGSSIASVRDLNQSYNDV 560
DB 1 GAPMYFNEGRERGVVYVELRQNFVYNGTLDKSHSYQNAFPGSSIASVRDLNQSYNDV 60
QY 561 VVGAPLEDNHAGAIYIFHGFGRSILKTPKORITASELATGLQYFGCSIHGQDLNEDGLI 620
DB 61 VVGAPLEDSHAGAIYIFHGFQTNILKVPQVQISASELAPGLQHFGCSIHGQDLNEDGLV 120
QY 621 DLAVGALGNVILWSPVQVQINASHLHFEPSKINIFHRDCKSRGDRATCLAAFLCFTPIFL 680
DB 121 DLAVGALGNVILWSPVQVQINASHLHFEPSKINIFHRDCKSRGDRATCLAAFLCFTPIFL 180

QY 681 APHFQTTVIGIRYNATMDERRVTPRAHLDEGGDRFTNRAVLLSSGOELCERINFHYLDTA 740
DB 181 APHFHTATVIGIRYNATMDERRVTPRAHLDEGGDRFTNRAVLLSSGOELCERINFHYLDTA 240
QY 741 DVVKPVTRESVEYSLSDPDHGPMLDGGWPTTLRVSVYFVWNGCNEDEHCVPLVLDARSIDL 800
DB 241 DVVKPVAESVEYSLSDPDHGPMLDGGWPTTLRVSVYFVWNGCNEDEHCVPLVLDARSIDL 300
QY 801 TAMEYCORVLRKPAQCSAYTSLFDTTFFIESTRQVAVEATLENRGENAYSTVLNISQ 860
DB 301 TAMEYCOVLRKPAQCSAYTSLFDTTFFIESTRQVAVEATLENRGENAYSTVLNISQ 360
QY 861 SANLOFASLIQKEDSDSGSIECVNEERRLOKQVCNVSYFFRAKAKVAFPLDSEFSKSI 920
DB 361 SENLOFASLIQKEDSDSGSIECVNEERRLOKQVCNVSYFFRAKAKVAFPLDSEFSKSI 980
QY 921 HLELELAGSDNSRSDSTKEDNVAPLPHLYEADVLFTRSSLSHYEVKLNSSLERVD 980
DB 421 HHLQIHLGAGSDSHSQSDTADNTALLRPHLYEADVLFTRSSLSHYEVKLNSSLERVD 480
QY 981 GIGPFPSCIFRIQNLGLFPIHGMWMMKIPIATRSNGNLLKRLDFLTDEANTSCNIWNS 1040
DB 481 GIGPFPNCVFKVQNLGFFPIHGMWMMKIPIATRSNGNLLKRLDFLTDEANTSCNIWNS 540
QY 1041 TEYRTPVEEDLRRAPQNLHNSDVSVCNIRLVNPOEINPHLLGNLWLSLKALKYKS 1100
DB 541 TEYRSTPTEEDLSHAPQNHNSDVSVCNIRLVNPOEINPHLLGNLWLSLKALKYKS 600
QY 1101 MKIMYNAALQRFHSPFIREDPSRQLEFEISKOEDWQVPIIIVGSLGLLALLIV 1160
DB 601 MKITVNAALQRFHSPFIREDPSRQLEFEISKOEDWQVPIIIVGSLGLLALLIV 660
QY 1161 LALXKLGFRRSARRRREFGLDTPKYLE 1188
DB 661 LALXKLGFRRSARRRREFGLDTPKYLE 688

RESULT 12

US-09-866-050A-501
; Sequence 501, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:

; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-501

Query Match 50.0%; Score 3113; DB 10; Length 696;
Best Local Similarity 85.3%; Pred. No. 7.5e-278;
Matches 584; Conservative 45; Mismatches 56; Indels 0; Gaps 0;

QY 501 GAPMYFNEGRERGVVYVELRQNFVYNGTLDKSHSYQNAFPGSSIASVRDLNQSYNDV 560
DB 1 GAPMYFNEGRERGVVYVELRQNFVYNGTLDKSHSYQNAFPGSSIASVRDLNQSYNDV 60
QY 561 VVGAPLEDNHAGAIYIFHGFGRSILKTPKORITASELATGLQYFGCSIHGQDLNEDGLI 620
DB 61 VVGAPLEDSHAGAIYIFHGFQTNILKVPQVQISASELAPGLQHFGCSIHGQDLNEDGLV 120

```
QY 621 DLAVGALGNVILWRPVVQINASLHFEPKINIFHRDCKRSGRDATCLAAFLCFTPIPL 680
Db 121 DLAVGALGNVILWRPVVQINASLHFEPKINIFHKDCKRGRDATICLAAFLCFTPIPL 180
QY 681 APHFQTTTGIIRYNATMDERRYTTPRAHLDGGDRFTNRAVLSSGQELCERINFEHVLDTA 740
Db 181 APHFQTTTGIIRYNATMDERRYNPRAHLDGGADQFTNRAVLSSGQEHQCORINFEHVLDTA 240
QY 741 DYVKPVTFSVEYSLEDEHGMDDGWPFTTLRLVSVFPWNGCNEDEHCVPLDVLDAESDLP 800
Db 241 DYVKPVPFVSVEYSLEDEHGMPLDNGWPFTTLRLVSVFPWNGCNEDEHCVPLDVLDAESDLP 300
QY 801 TAMEYCORVLRKPAQDCSAYTISFDVTTFIIESTRORVAVATLENRGENAYSTVLNISO 860
Db 301 TAMEYCCQVLRKPAQDCSSYTISFDVTTFIIESTRERRVAVATLENRGENAYSAVLNISO 360
QY 861 SANLQFASLIQKEDSGSTECVNEERRLOKVCNVSPFPRAKAKVAFRLDSEFSSIFL 920
Db 361 SENLQFASLIQKODSINSIECVNEERRLKKVCNVSPFPRAKAKVAFRLDSEFSSVFL 420
QY 921 HHLEIELAAGSDSNRSDKEDNVAPLRFHLKYEADVLFTRSSLSHVEVKLNSSLERYD 980
Db 421 HHLQIHLGAGSDSHEQDSTADDNTALLRFHLKYEADVLFTRSSLSHFEVKANSLESYD 480
QY 981 GIGPPSCIPRIQNLGFLPHGMMKTIPIATRSNGNRLILKLDFTLDEANTSCNIWGS 1040
Db 481 GIGPPENCVPKQNLGFFPHGVMKTIPIATRSNGNRLMLKDFDFTDQVNTSCNIWGS 540
QY 1041 TERYPTPVRDLARAPOLNHSNDVSVINCINVLVNOEINPHLLGNLWLSLKALKYKS 1100
Db 541 TERYPTTEEDLHAPORNHSNDVSVIICNVRLAPNQETSFYLVGNLWLSLKALKYKS 600
QY 1101 MKIMVNAALQRFHSPFIREEPSRQIEFEISKOEDWQVPIIIVGSTLGGLLALLV 1160
Db 601 MKITVNAALQRFHSPFIREEPSRQVTFEISKOEDWQVPIIIVGSTLGGLLALLV 660
QY 1161 LALRKLGFERSARRRREPGLDTPK 1185
Db 661 LALGSLVSLKVPARGSPAWAPSPK 685
```

RESULT 13

```
US-09-866-050A-500
; Sequence 500, Application US/09866050A
; Publication No. US2003004071A1
```

GENERAL INFORMATION:

```
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Marison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011C4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 500
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Mouse
```

US-09-866-050A-500

```
Query Match 40.5%; Score 2518; DB 10; Length 545;
Best Local Similarity 86.2%; Pred. No. 4.8e-223;
Matches 470; Conservative 35; Mismatches 40; Indels 0; Gaps 0;
```

```
QY 644 SLHFEPKINIFHRDCKRSGRDATCLAAFLCFTPIFLAPHQTTTGVIRYNATMDERRYT 703
Db 1 SLHFEPKINIFHKDCKRGRDATICLAAFLCFTPIFLAPHPTATVIRYNATMDERRYM 60
```

```
QY 704 PRAHLDGGDRFTNRAVLSSGQELCERINFEHVLDTADYVKPVTFSVEYSLEDEPHGPML 763
Db 61 PRAHLDGGADQDFTNRAVLSSGQEHQCORINFEHVLDTADYVKPVAFSVEYSLEDEPHGPML 120
QY 764 DGGWPTTLRLVSVFPWNGCNEDEHCVPLDVLDAESDLP TAMEYCORVLRKPAQDCSAYTIS 823
Db 121 DNGWPTTLRLVSVFPWNGCNEDEHCVPLDVLDAESDLP TAMEYCCQVLRKPAQDCSSYTIS 180
QY 824 FDTTFVFIIESTRORVAVATLENRGENAYSTVLNISOANLQFASLIQKEDSDGSTEVCN 883
Db 181 FDTTFVFIIESTRRRVAVATLENRGENAYSAVLNISOENLQFASLIQKDDSDNSTEVCN 240
QY 884 EERRLOKVCNVSPFPRAKAKVAFRLDSEFSSIFLHLEIELAAGSDSNRSDKEDN 943
Db 241 EERRLHKVCNVSPFPRAKAKVAFRLDSEFSSVFLHHLQIHLGAGSDSHEQDSTADDN 300
QY 944 VAPLRFHLKYEADVLFTRSSLSHVEVKLNSSLERYDVGPPFSCIFRIONLGLPFIHGM 1003
Db 301 TALLRFHLKYEADVLFTRSSLSHFEVKANSLESYDVGIPFENCVPKQNLGFFPHGV 360
QY 1004 MKKTIPIATRSNGNRLILKLDFTLDEANTSCNIWGNSTERYPTPVRDLARAPOLNHSN 1063
Db 361 MKKTIPIATRSNGNRLMLKDFDFTDQVNTSCNIWGNSTERYPTTEEDLHAPORNHSN 420
QY 1064 DVSVINCINVLVNOEINPHLLGNLWLSLKALKYKSMKIMVNAALQRFHSPFIREE 1123
Db 421 DVSVIICNVRLAPNQETSFYLVGNLWLSLKALKYRSMKITVNAALQRFHSPFIREE 480
QY 1124 PSRQIEFEISKOEDWQVPIIIVGSTLGGLLALLVLAALRKLGFERSARRRREPGLDPT 1183
Db 481 PSRQVTFEISKOEDWQVPIIIVGSTLGGLLALLVLAALRKLGFERSARRRREPGLDPT 540
QY 1184 PKVLE 1188
Db 541 PRELE 545
```

RESULT 14

```
US-10-741-601-531
; Sequence 531, Application US/10741601
; Publication No. US20040166519A1
```

GENERAL INFORMATION:

```
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 531
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-10-741-601-531

```
Query Match 38.8%; Score 2412.5; DB 16; Length 1167;
Best Local Similarity 43.0%; Pred. No. 1e-212;
Matches 514; Conservative 213; Mismatches 424; Indels 45; Gaps 16;
```

```
QY 1 MDLPRGLVAVAWALSLWPGFTDTTFNMMDTRKPRVIPGSRTPAFFYTVQOHDISGNKWLVGA 60
Db 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLPFGPPEAFEGYSVLQHVGGQRWMLVGA 60
QY 61 PLETNGYQKTDGVYKQPV----IHGNC TKLNLGRVTLNSVNRKDNMPLGLSLATNPKNS 117
Db 61 PWDGSDRRRQDGYRCPVPGGAHNAKAGHLG DYQLGNSSHPAVNMLHGLMSLLETDTGGG 120
QY 118 FLACSPWLSHECCSSYYVTTCGRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Db 121 FMACAPLWSRACSSVFESSGICARVDASFQSGSLATPQRCPTVMDVIVLDGNSIYP 180
```


Db 780 QKLVFSPKDCGPDNECVTLVLQVNDI-----RGRK-----APFVV 817

QY 832 ESTQRVAVEATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSIKCVNEERLQKQ 891

Db 818 RGRKRVLVSTTLENKENAYNTSLIFSRNLHLASLTPQRESPIKVECAAPS--AHAR 875

QY 892 VCNVSYPPFFRAKAKVAFRLDSEFSKSIPLHLEIELAAGSDSNERDSTKEDNVAPLRFHL 951

Db 876 LCSVGHEVFTGAKVTFLEFEFSCSLLSQVEFKLTASSDSLERNGLQDNTAQTSAVI 935

QY 952 KYEADVLFTRSSLSLHYEVKINSLEKYDYGPPFSCIPRIQNLGLFPFHGMMKITIPI 1011

Db 936 QYEPHLLFSESITHRYEYHPYGTLP--VGPGEFKTTLRVQNLGCYVVSGLIISALLPA 993

QY 1012 ATRSGNRLKLRLDPLTDEANTSCNWNSTYRPTVE-EDLRRAPQLNHSNDVVSINC 1070

Db 994 VAHGNVFLSLSQVITN--NASC-IVQNLTEPPGPPVHPEELQHTNRLNGSNTCCQVRC 1050

QY 1071 NI-RLVFNQEIHFHLLGNLWRLSLKALKYKSMKIMVNAALQRFHSPFFIFREEDPSRQIE 1129

Db 1051 HLGQLAKGTEVSVGLLRLVHNEFFRAKFKSLTVSTFELGTBEGSVLQJTEASRWSESL 1110

QY 1130 FEISKQEDWQVPIWIVGSTLGGLLIALILVLAIRKLGFFRSAR----RRREPGLD 1181

Db 1111 LEVVQTRPILISLWILLIGSVLGGLLIALLVFCLWKLGFPAHKKIPBEEKREEKLE 1166

Search completed: September 21, 2004, 13:29:01
Job time : 219.039 secs


```
QY 71 GDVYKCPVHGN---CTKLNGL-RVTLNVSERKDNMRLGLSLATNPKNDSFLACPLWS 126
Db 79 GDVYKCPVGRGESLFCVLDLDPVNTSIPNVTEVKENMTFGSLVTNP-NGGSLACGLPYA 137
QY 127 HECSSYTTGMCSSVNSFRFSKTVAPALQRCQYMDIVIVLDGNSNSIYPVWEQHFLL 186
Db 138 YRCGLHLHTTGCSVDVPTFQVNSIAP-VQCSYXLDIVIVLDGNSNSIYPMDST-ALN 195
QY 187 NILKFPYIGPQIQGVQVQYGVVHFEHLNDYRSVKDVEAASHIEQRGGTETRTAFGI 246
Db 196 DLLKRMIDGPKXTXGVIVXGENVTHEFNLNKYSTEEVLVAAKKIVXRGGRXTWTALGT 255
QY 247 EFARSEAP--QXGGRGAKKVMIVITDGBSHDPSPLEKVIQOSERDNTVRYAVAVLGYN 304
Db 256 DPARKEAFTEARGARGVKKVMIVITDGBHDXNHLKKVIGCBEDENXRFSAIILGSYN 315
QY 305 RRGINPEFLNIEKIYASDPDDKHFNVTDEAALDIVDALGRIFSLGTT-KNNETSFG 363
Db 316 RGNLSTEKEFEVEIKSIASEPTKSPFNVSDELALVTIVKTLGERIFALEATADXAASFE 375
QY 364 LMSQTGSSHVVEGVLLGAVGAYDNGAVLKETSAGKVIPLRBSYKKEFPEELKNHGA 423
Db 376 MEMSQTGSAHYSQDMVWMLGAVGAYDNGVTNVNQASQIIIPRNTTFNVESTKKNEPLAS 435
QY 424 YLGYTVTSVVSRRQGVYVAGAPRNHTGKVLFTMHNNRSLTIHQAMRGQOIGSYFGSE 483
Db 436 YLGYTVNSATASGVDLYTAGFRYNTGQVIYRMEEG-NIKILQTLGSGXIGSYFGSI 494
QY 484 ITSVDIDGQVTDVLLVGPMTYF-NEGRERGKVYVVELRQNRPFVYNGTL----- 531
Db 495 LITTDIDKDSNTNILLVGPMTYGTKEBGGKVYVVALNQTPEYQMSLAPMEPIKOTCC 554
QY 532 ----KDSHSYQN-----ARFGSSIASVRLNQDSYNDVVGVGAPLEDNHAGAIYIFHGR 581
Db 555 SSRQHSCTTENKNEPCGARFQTAIAVKDLNLDGNDIVIGAPLEDHGGAVYVHSGS 614
QY 582 GSILTKPKQRIATASELATGLQVFGSIHQGLDINEDGLIDLAVGALNAVILWSRPVQI 641
Db 615 KTRKEYAARIFSGGDKTLKFGQSIHGEMDLNGDLTDVTCGLGGAAFLWSRDVAV 674
QY 642 NASLHPEPSKINIFHRDCKRSQRDATCLAAFLCFTPIFLAPHFTTIVGIRYNATWDER 701
Db 675 KVTMAFEPKNVLOKNCHEGKETVCINATVCFEVKLASKEDTIYEADLQYRVILDSLR 734
QY 702 YTPRAHLDGGRFTNRNAVLLSGGBELCERINPHVLDTADYVKPVTFSVEYSLEDDHGP 761
Db 735 QISRSFFSGTOERKQVNTIVRKSE--CTKHSFYMLDKHDFQDSVRITLDFNLTPENG 792
QY 762 MLDDGWPTTLRVSVPPWNGCNEDEHCVPLDVLDAKSDLPATAMEYCORVLRKPAQCSAYT 821
Db 793 VLDDSLPNSVHYIIFPAKDCGNKEKISDL-----SLHV 826
QY 822 LSFDTTFTIESTROKVAVEATLENRGENAYSTVLNISQASLQFASL--IQKEDSDGSI 879
Db 827 ATTERDOLLVRQNDKFNVSILVTKNTKDSAYNTRITVHVSFNLVFSGIEAIQKDSCESN- 885
QY 880 ECVNEERLQKQVNCVSYPPFRAKAVAPRLDSEFSKSIPLHLETELAAAGSDSNERDST 939
Db 886 -----HNITCKVGYPLRGGEMVTEPKILFQENTSYLMENVTIYLSATSDEBPET 936
QY 940 KEDNAPLRLHLYEADVLFTSSLSLHVEVKLNSS-----LERYDGIQPPFSCIFRIONL 995
Db 937 LSDNVNVSIPVUYEVLGYFYSASBYHSIAANETVPEVINSTEDIGNEINIFYLRKS 996
QY 996 GLFPIHGMKMTIPIATRSNRLRLKLRDLFTDEANTSC-----NINWGNSTEYRTPPVE 1049
Db 997 GSPFMPBLKLSIFPNNTSGYPVL-VPTGLSSSENANCRPHIFEDFPFSGINSKMTTST 1055
QY 1050 EDLRAPQLNHSNDVVSINCLRLVFNQEIHFLLGNLWLSLXALKYKSKMIMVNAAL 1109
Db 1056 DHLKRGITLDCNTCKPATITCNLTSSDISXANVSLI--LWKPTFKSYFSSNLNLTIRGEL 1113

1110 QROFHSPIFREEDPSROIEFEISKOE-DWOVPIWIVGSTIGLLALLLALLVLRKLG 1168
1114 -RSENASVLSSNKKRELAIQISKDGLPGRVPLWLLSAPAGULLMLLILALWKIGF 1172
1169 FRSARRR 1176
1173 FKRPLKK 1180

RESULT 2
US-09-532-310B-6
; Sequence 6, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
; ; Detmar, Michael
; ; Claffey, Kevin P
; TITLE OF INVENTION: Method for inhibiting tumor
; ; angiogenesis in a living subject
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, Esq.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: B1S-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-532-310B-6

Query Match 29.1%; Score 1813.5; DB 4; Length 1183;
Best Local Similarity 35.1%; Pred. No. 2.4e-154;
Matches 428; Conservative 239; Mismatches 434; Indels 117; Gaps 33;

QY 4 PRLGVVALSLWPGFTD---TFNMDTKPRVIPGSRATFAFGYTVQOHDISGNKMLVGA 60
Db 10 PLPLLLVLALS--QGLNCCCLAYNGLPEAKTFSPSSSEQFYAVQOFTNPKGNLLVGS 67
QY 61 PLETNGYQKTGDVYKCPV--IHGNTCTKLNGLRVT--LSNVSEKDNMRLGLSLATNPKNDS 117
Db 68 PMSGFPENRGDVKCPVDLSTATCEKLNQSTSIPTNVTEKTNMSLGLILTRNMGTS 127
QY 118 FLACPLWSHECGSSYYTTCMSRVNSNFRFSKTVAPALQRCQYMDIVIVLDGNSIYP 177
Db 128 FLTCTGPLMAQCGNQYTTGVCSDISDPDFKXSAFSPATXPCPSLIDVWVWVCDSESIYP 187
QY 178 WVEVQHFLLNLIKFKYVGPQIQGVQVQYGVVHFEHLNDYRSVKDVEAASHIEQRGG 237
Db 188 WDAVKNFLKFKVQGLDIPGTTQVGLIQVANNPRVVFNLTYYTKREMIATVTSQYGG 247
QY 238 TETRTAFGIEFARSEAFQ--KGRKGAKKVMIVITDGBSHDPSPLEKVIQOSERDNTVRY 295
```

Db 248 DLTNFGALQYARKYAYSAAGRESATKVMVVVTDGSHDGLMKAVIDQCNDHNIIRP 307
QY 296 AVAVLYNNRGINDE-TFLNEIKIADSPDDKHPFNVTDEAALKDVIDALGDRIFSLGEG 354
Db 308 GIAVLGLNNRNDLPMNLIKAIKAIASIPTERYFNFVDEAALIEKAGTLGEQIFSTEG 367
QY 355 TNKNETSFGLEMSQTFSSHV--VEDGVLLGAVGAYDNGAVLKETSAGKVIPLRESYLK 412
Db 368 TVXGDNFXEMXSQVGFADYSSQNDILMLGAVGAFGWSGTIVQKTSIHGHLI----- 419
QY 413 EPPEEL-----KNHGAYLGYTVTSVWSRQGRVYVAGAPRFNHTGKVLFTMNNRSLT 466
Db 420 -FPKQAFQIQDRNHSSYLGSV-AAISTGESTHEVAGAPRANVTGQIVLYSVNENGNI 477
QY 467 IQHMRGQIGSYFGESETSDVIDDGGVTDVLLVGCAPMYNE-GRERKGVVVELQRNRF 525
Db 478 VIQHRGQIGSYFGEVLCSDVDKDTITDVLVGPAPMYMSDLKKEGRVYLFITIKGIL 537
QY 526 VYNGTLKDSHVQNARFGSSIASVRDLNODSVNDVVVCGAPLEDNHAGAIYIFHGFGRGIL 585
Db 538 QHQFLEGEPEGIEVTRFSAALASDINDGFDNDVIVGSPLENQSGAVIYNGHQGTIR 597
QY 586 KTKPQRITASELA--TGLQYFCSTHGOLDNEDGLIDLAVGALGNVILMSRPVWQINA 643
Db 598 TKYSQKILGSDGAFRSHLQYFGRSLDGYGLNGDSITDVSIGAFQVQVQLWSQSIADVAI 657
QY 644 SLHEPESKINIFHRDKCRSGRDATAFLAFCFTPIFLAPHQTTTGVGRYNATMD-----E 699
Db 658 EASTPEKITIVNNAQ-----IILKLCFSAKF-RPTKQNNQVAIVYNTILDADGFS 708
QY 700 RRYTPRAHLDRGGDRFTNRAVLGSGBELCRINFHVLDTADYKPV-TFSVEYSLEDPD 758
Db 709 SRVTSRGLFKENNERCLOKMNVMVQAQCPHEI-IYOEPSDVNSLDLVDVLSLENGP 767
QY 759 HGPMLDDGWPTTLRVSVFPMWNGCNEDEHCVPLVL-DARSLDPTAMEYCORVLRKPAQDC 817
Db 768 TSPAELAYSEYAKVFSIPFHKDCGEDGKCIDSLVLQDVR--LPAQE-----QP----- 814
QY 818 SAYTSLFDTTFTIESTRQVAVATLENRGENAYSTVLNISQSANLOFASLQKEDSDG 877
Db 815 -----FIVSNQNKRLTFSVTLKXKRESAYNTGIIVDFSENLFASFSLPVD--- 860
QY 878 SIECVNERRIKQOV-CNVSVYFPRAKAVFRLOSEFSKIFULHLEIELAAGSDSNR 936
Db 861 GTEVTCQVAASQKSVACDVGPALKREQVITFINDFNLQNLQNASLSFQALUSESEE 920
QY 937 DSTKEDNVAPLRFHLKYEADVLFTRSSLSHYEV-----KLNSSLEYDYGIGPPFSCIPRI 992
Db 921 N--KADNLVNLKIFLLYDAETHLSTRNINFEISSDGNVFSIVHSPEDVGPKF--IFSL 976
QY 993 Q-NLGLPFIHGMKTIPIATRSNRLKLRLDPLTDEANTSCNIWGNSTYRTPV--- 1048
Db 977 KVTGSPVSVMATVILHLPQTKENPLMYITGVTQDKAGDIC-----CNADINPLKIGQT 1032
QY 1049 -----EEDLRAPOLNHSNDVVSINCNRLVFNQBINHLG-----NLWLRSL 1093
Db 1033 SSSVSFSKSNRPHKELNCRPTASCNSVTCWK-----DVHMKGEYFVNTTRINWGT 1085
QY 1094 KALKYKMKIMVNAALQRPSPFIIFREDEPSRQIEFEISK-QEDWQVPIIIVGSTLGG 1152
Db 1086 ASSTFQT--VQLTAAAEINTYNEPIYVIEDNTVTIFLIMKPEDEKAEVPTGVIIGSIAG 1143
QY 1153 LILLALLVIALRKLGFFR 1170
Db 1144 ILLALLVAILWKLGFEX 1161

RESULT 3

US-08-485-618-53
; Sequence 53, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-53

Query Match 19.2%; Score 1194.5; DB 1; Length 1161;
Best Local Similarity 30.1%; Pred. No. 2.3e-98;
Matches 374; Conservative 198; Mismatches 507; Indels 165; Gaps 49;

QY 5 RGLVV---ANALSLWPGFTDTFNMDTKPRVPGSRTPAFGYTYQOHDISGNKMLVYCAP 61
Db 3 RGVVILLCGWALASCHG-----SNLDVEKP-VYFKEDAASFGQTVVQF---GGSRLVWGAP 54
QY 62 LETNGYOKTGDVYKCPVIHGNCTKLNLRVTLNSVSRKXNMRLGLSLATNPKDNSFLAC 121
Db 55 LEAVAVNQTCOSSDCPATGVCQPIILL-HIPLEAV-----NMSGLSLVADTNNSQLLAC 108
QY 122 SPLMHSCHGSSYYTTCMCSRVSNNFRFSKTVAPALQRCQ-TYMDIVIVLDGNSNI--YPM 178
Db 109 GPTAQACAKMNAKAGSCLLIGSLQFIQAIPTATMPCPGQEMDIAFLIDGSGSIDQSDP 168
QY 179 VEVOHFILNLIKFEYIGQGI-----QVGVVQGEDVVHFEHLNDYSKVDKVEAAASHIE 233
Db 169 TQMKDFVKALM-----GQLASTSTSFSLMYSNLIKLTHTFTTEFKSSLSPOSLSLVDAIV 221
QY 234 QRGGTETRTAFGIEFARSEAF--QKGQRKGAKKVMIVITDGESHDS-DLEKVIQOOSERD 290
Db 222 QLQGL-TYTASGQKQVKVKELFHSGNKGARKSAKKILIVITDQKPRDPLETRHVIPEAKA 280
QY 291 NVTRYAVAVLYGYNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAALKDVIDALGDRIF 350
Db 281 GIIRYAIGVDAPFE-----PTALQELNTIGSAPSQDHFVKGVNFVALRSIQRQIQEKIF 335


```

Db 449 EVRGTOIGSYFGASLCSVDMDDRGDSTDLVLIGVPHYEHTR--GGQVSVCPMPGVRSRWHC 507
Qy 528 NGTLKDSHSYQNAFSGSSIASVRDLNQSDYNDVVVGALENDHAGAIYIFHG--FRGSILK 586
Db 508 GTTLHGEGHPWGRFGAALTVLGDVNGSDADVAIGAFGEENRGAVYIFHGASRQDIAP 567
Qy 587 TPQRITASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSRPVQVQNASLH 646
Db 568 SPSQRTVGSQFLRLQYFGSLGQDITQGLVDLAVGAQHVLRLSLRLEKRGVIR 627
Qy 647 PEPKINIFHRDKCRSGRDATACL---AAFLCFT-----PIFLAPHFQTTTIGIRYNATMD 698
Db 628 FAPSEVAKTVQC--WGRTPVLEAGATVCLTVRKSGPDLLG---DVQSSVRYDLALD 681
Qy 699 ERRYTPRAHLEGGDRFTRNAVILSSGOELCERINFHVLDTADYVVKVPTFSVEYSL---754
Db 682 PGRUISRAIFDETKNCTLRKTKTGLGDH--CETMKLLPDCVEDAVTPIILRLNLSAGD 740
Qy 755 BDPDHG--PMLDDGWFTTLRVSVFVWNGCNEDEHCVPLDVLDAKSDLPAMEYQORVLRK 812
Db 741 SAPSENLRPLVAVSGSDHVTASFPEKNCKQELLCEGNL-----779
Qy 813 PAQCSAYLTSFDTFTVFIETSTORVAVEATLENGENAYSTVLNLSQSANLQFASLTQK 872
Db 780 -----GVSENFSGQLQVLEVGSSPELTVTVTWNEGEDSYGTLIKFYYPALSYRVRTA 833
Qy 873 EDSDG---SIECVNE---ERSLQOVQNVSVPPFRAKAKAVAFRLDSEFSKIFL--HHLEI 925
Db 834 QQPHYPILRJACEAPTQCESLRSSCSINHPIFREGAKATFMTFDVSYKAFGLDRLL 893
Qy 926 ELAAGSDSNERSDKEDNVAPLRHLKYEADVLTFRSS-----SLSHVEVKINSLE 977
Db 894 RASASSENKETS--TAFQLELPVKYTVTVVISRQEDSKHFNFSHGE--RQKRAH 950
Qy 978 RYDGIQPPFCIFRIONGLFPIHGMKMTIPIATRSKNL--LKLDRFLTDENATSCN 1035
Db 951 RY-----RVNLSPLTL--AISVNFVFPILL--NGVAVMDVTLR---SPAQGVSC- 993
Qy 1036 IWGNSTEYRTPVEEDLRAPO---LNHSNSDVVSINCN--RLVPQOEIFHLLGNL--1088
Db 994 -----VSGREPQSHDLLTQIGRSVLDCAIDCLHLECDIPSLGTDLDELFIKGNLSF 1048
Qy 1089 -WLRLS---KALKYKMKIMVNAALQRFHSPFFIFREEDPSRQLEFEISKQEDQVPIWI 1144
Db 1049 GWISQTLQKKVLLSEAEITNTSVYSQLPQEAFLRAQVNTMLEEYVYV--PVL 1103
Qy 1145 IVGSLGGLLLALLVIALRKLGRFSARRRE---PGLDPTP 1184
Db 1104 MVFSSVGGLLLLALITVALYKLGFFK--RQYKEMLDLPSADPDP 1145

```

RESULT 5

```

US-08-605-672-53
; Sequence 53, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515e1 Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; STREET: Marshall, O'Toole, Gerstein, Murray & Borun
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-605-672-53

```

```

Query Match 19.2%; Score 1194.5; DB 2; Length 1161;
Best Local Similarity 30.1%; Pred. No. 2.3e-98;
Matches 374; Conservative 198; Mismatches 507; Indels 165; Gaps 49;

Qy 5 RGLVV---AWALSMPGFTDTFNMDTKRPVPGSRITAFGYTVOQHDISGNKWLWVGAP 61
Db 3 RGVVILLCGWALASCHG---SNLDVEKP--VVFVEDAASFQTVVQF---GGSRLVVGAP 54
Qy 62 LETNGYKTKGDVYKCPVHGNCTKLNLRVTLSNVSRKONMRGLSLATNPKNSFLAC 121
Db 55 LEAVAVNQTGSSDCPPATGVCQPILL--HIPLEAV-----NMSGLSLVADTNNSQLLAC 108
Qy 122 SPLWSHSCGSYYTTCMSRVNSNFRSKTVAPALQRCQ--TYMDIVTVLDGNSI--VPW 178
Db 109 GPTRAQACAKMAYKAGSCLLLGSSLOFQIAPATMPCPGQEMDIAPLIDGSGIDQSD 168
Qy 179 VEVOHFILNLIKFYIGGQI---QGVVQYGEDVHVHFNLDYRSKDVVWEAASHIE 233
Db 169 TQMKDFVKALM-----GQLASTSTFSLMQYSNLIKTHFTTFEKKSSLSPOSLVDAIV 221
Qy 234 QRGCTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHDSP--DLEKVIQOBERD 290
Db 222 QLQGL--TYTASGIQKVVKELFHSKNGARKSAKTLIVITDGQFRDPLEYRHHVPEAKA 280
Qy 291 NVTRYAVAVLGYNNRRGINPETFLNEIKYIASDDDDKHFFNVTDAAALKDIDVADGRIF 350
Db 281 GILRYAIGVDAFRE-----PTALQELNTIGSAPSDHVFVKVGNFVALRGIQRIQEKIF 335
Qy 351 SLEGT--KNKETSFGLEMSQTGFSHVVDGVLGAVGAYDWNGAVLKAVLAKETSGAKVILRES 409
Db 336 AIEGTESRSSSSSQHEMSQEGFSALSMDGPFVLGAVGFSWGGAFLYPS-----NWRST 390
Qy 410 YLKEFPEELKHGAYLGTYTSSVVSQGVYVAGAPRPHHTGKVLILFTMNNRSLTIHQ 469
Db 391 FINNQENEDMRDAYLGS--TALAFWKGVHSLILGAPRHQHTGKVIIFT--QESRHWPKS 448
Qy 470 AMRQQIGSYFGSEITTSVDIDGVDGVTDLVLVGAPMYFNEGRGVKVVVYELR--ONRFVY 527
Db 449 EVRGTOIGSYFGASLCSVDMDDRGDSTDLVLIGVPHYEHTR--GGQVSVCPMPGVRSRWHC 507
Qy 528 NGTLKDSHSYQNAFSGSSIASVRDLNQSDYNDVVVGALENDHAGAIYIFHG--FRGSILK 586
Db 508 GTTLHGEGHPWGRFGAALTVLGDVNGSDADVAIGAFGEENRGAVYIFHGASRQDIAP 567
Qy 587 TPQRITASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSRPVQVQNASLH 646

```

Db	568	SPSQRTGSQLFRLQYFGQSLGGQDLTQDGLVLAAGQGHVLLRLSRPLIKVGISR	627
QY	647	FESKINIPHRDCKRSRDATCL--AAFLCFT----PIFLAPHFOTTVGIRYNATMD	698
Db	628	FASEVAKTVYQC--WGRTFVLEAGEATVCLIVRKGPSDLLG----DVQSSVRYDLALD	681
QY	699	ERYTTPRAHLDGGDRFTNRAVLLSSQOELCERINFHVL-DADYKVPVTFVSVEYSI---	754
Db	682	PGRLLISRAIPDETKNCTLTRKKTIGLGDH-CETMKLLLPDCVEDAVTPIILRLNLSAGD	740
QY	755	EDPDHG--PMLDQGWPTTLVSVSPFMWGNEDHCVLDLVARSDLPTAMEYCQRVLRK	812
Db	741	SASRNLRPVLAGSQDHVITASPFCKCKQKQELLCEGNL-----	779
QY	813	PAQDCSAYTLSPDTTVFIIESTQRVAVEATLENRGENAYSTVNLINSQSANLOFASLIQK	872
Db	780	-----GVSPNFSGLQVLEYGSSPELTVTVTWNNEGDSYGTLLKFFYPAELSYRRVTRA	833
QY	873	EDSDG---STECVNE---ERRLOQVCNVSPYPPFRAKAKVAFRLDSFSKSIFL-HHEI	925
Db	834	QQHPYPLRIACAEAPFGQBSLRSSCSINHPFREGAKATMTIPDVSKYAFGLDELLL	893
QY	926	ELAAGSDNSRSDTKEDNVAPLPHLKYEADVLTFRSS-----SLSHYEVKLNSLE	977
Db	894	RASASENNKPETSK--TAFGLELPVKYTVTVISROEDSTKHFNFSSHGE-RQKEAEH	950
QY	978	RYDIGIGPFPSCIFRIONLGLFPIHGMKMITIPIATRSGNRL--LKLRDFTDEANTSCN	1035
Db	951	RY-----RVNMLSPILT--AISVWFVPILL-NGVAVMDVTLR----SPAQGVSC-	993
QY	1036	IWGNSTBYRTPVEEDLRRAPQ-----LHNSNSDVVSINCN-RIVNPQCEINFHLGNL--	1088
Db	994	-----VSQREFPQHSDDLTOQGRSVLDCAIADCLHRCIDPSLGTLDLDFILKGNLSF	1048
QY	1089	-WLRLS---KALKYKSMKIMVNAALQRFHSPPIFREEDPSRQTEFEISKOEDQVPIWI	1144
Db	1049	GWISQTLQKKVLLLSAEITFTNSVYSQLPGQEAFLRAQVSTMLEEFVVE-----PVFL	1103
QY	1145	IVGSTLGLLLALLVILAKRLKGFRRARRRE-----PGLDPTP	1184
Db	1104	MVFSVSGVGLLLALITVALYKLGFFK--ROYKEMLDLPSADPDP	1145

RESULT 6

291	NVTRTAVAVLGYNNRGINPETFLNEIKYIASDDDDXHFNNVTDEAALKOIVDALGRDIF	350
281	GIIRYAIGVGDAFRE-----FTALQELNTIGSAPSDHVFVKGVFNVALRSIQROIQEKIF	335
351	SLEGT-NKNETSFGLMSQTGTFSSHVEDGVLLGAVGAYDWNGAVLKETSAGKVIPIRES	409
336	ALEGTESRSSSSSQHEMSQEGFSSALSMDGPVLGAVGFGFSWSSGAFLYPS-----NMKST	390
410	YLKEFPEELKXHGAYLGYTVTSVSSRQGRVVVAGAPRFNHTKVLFTMHNNRSLATHQ	469
391	FINMSQENEDMDAYLGVS-TALAPWKGVHSLILGAPRHQHTKGVIFT-QESHWPEKS	448
470	AMRGOQIGSYFGSEITSDVDGDGVTDLVLLGAPMYFNREGRGKVVYIELR--QNRFPVY	527
449	EVRGTTIGSYFGASLCSVMDRDGSTDVLVIGVPHYBHTR-GQGVSVCPMPGVRSWHC	507
528	NGTLKDSHSYQNAFSGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIHG-FRGSILK	586
508	G'TLHGEGHPWGRFGAALTVLGDVNGDSLADVAITGAPGEENRGAVVIFHGASRQDIAP	567
587	TPKQRTASELATGYFGCSITHGQDLNBERGLDLAVGALGNVALWSRPVVOINSLH	646
568	SPSRVGTGSQLFLRLQYFGQSLSGGQDLTQDGLVDLAVGAQGHVLLRSLSLKVGISIR	627
647	FEPSKINIFHRDCKSRGDRATCL---AAFLCFT-----PIFLAPHFTTTVGIRYNATMD	698
628	FAPSEVAKTVYQC--WGRTPTVLEAGEATVCLTVKRGSPDLLG-----DVQSVRYDLALD	681


```

QY      813 PAQDCSAYTTFDOTTFFIIBESTQRVAVEATLENRGENAYSTVLNTSQSANLQFASLIQK 872
Db      780 -----GVSFNFSGLQVLEVGSSPELTVTVVWNEGEDSYGLIKFYPAELSYRRVTRA 833
QY      873 EDSGD-----STECVNE-----ERRLOQVCNVSYPFPRAKAKVAFLRDLSEFSKSIFL-HHLEI 925
Db      834 QQHPHYPLRLACEAEPGQESLSSSSCSINHIFREGAKATMTITFDVSYKAFGLDRLLL 893
QY      926 ELAAGSDNSRDSSTKEDNVAPLPHLYKEADVLFTRSS-----SLSHYEVKLNSSLE 977
Db      894 RASASSENKNPETSK--TAFQLELPVKYTVYTVISROEDSTKHFNFSSSHGE-RQKEAEH 950
QY      978 RYDGIQFPFSCIRIQNLGLFPIHGMMKKITIPATBSGNEL--LKLRLDLTDEANTSCN 1035
Db      951 RY-----RVNLSPLTL-AISVNTFWPILL-NGVAVMDVTLR-----SPAQGVSC- 993
QY      1036 IWGNSTYRPTPVEEDLRAPOQ---LNHSNSDVWSINCNI-RLVPNQEIFNHLGLNL-- 1088
Db      994 -----VSQREPPQHSDDLTOIQGRSVLDCAIADCLHURCDIPSLGTLDELDFILKGNLSF 1048
QY      1089 -WLRSL----KALKYKSMKIMVNAALQRFHSPFIIFREEDPSROTEFISKQEDQWVPIMI 1144
Db      1049 GWISQTLQKKVLLLSAEITFNTSVYSQLPQGEAFILRAQVSTMLEEYVYVE----PVFL 1103
QY      1145 IVGSTGLGLLLLLALVLALRKLGFRRGARRRE-----PGLDPTP 1184
Db      1104 MVSFSSVGGLLLLALITVALYKLGFFK--ROVKEMLDLPSPADDP 1145

```

RESULT 8

US-09-193-043-53
; Sequence 53, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el H

		Query Match	19.2%; Score 1194.5; DB 3; Length 1161;
		Best Local Similarity	30.1%; Pred. No. 2.3e-98;
		Matches	374; Conservative 198; Mismatches 507; Indels 165; Gaps 49;
Qy	5	RGLVV---ANALSLWPGETDFNMDTKPRVPGSRTAFTGYTQQHDISGNKLVLVGAP	61
Dd	3	RGVILLICGWLAKCHG-----SNLDVEKP-VVFKEAASFGQTVVQF---GGSELVLVGAP	54
Qy	62	LETNGYQKTGDVYKCPVIHGNCITKLNLGRVPLSVNSVERKNMRLGLSLATNPKNCSFLAC	121
Dd	55	LEAVAVNCTGSSDCPPATGVQCPIIL-HIPLEAV-----NWSLGLSLVADTNNSQLLAC	108
Qy	122	SPLNSHECGGSYYTTGMCSRWNNSFRSKTVAPALQRCC-TYMDIVIVLDSGSI--YPW	178
Dd	109	GPTAQACAKNNYAKGSCLLIGLSLTQFIATPMTCPCPGEMDIAFLIDSGSIDSDF	168

Db 576 SLSRLQYFGQSLGGQDLTQDGLVDLAVGAQGHVLLLSRLPLLLKVELSIRFAPMEYAK 635
Qy 655 FHRDCKRGRDATCL---AFLCFTPIFLAPHFQTTTVG-IRYNATMDERRYTTPRAHLDE 710
Db 636 AVYQCWE--RTPTVLEAGEATVCLTVHKGSPDLLGNVQGSVRYDLALDPGRLLISRAIFDE 693
Qy 711 GGDRTNRAVLSSSQELCERINPHVLD-TADYVKPVTFSVEYSI-----EDPDHGMULD 764
Db 694 TKNCTLTGRKTLGLGDH-CETVKLLLPDCVEDAVSPFIILRLNFSLVRSASPRNLUHPVLA 752
Qy 765 DGMPTTLRVSPFMWGNCDNEHCVPDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTLSF 824
Db 753 VGSQDHIATSLPFKNCKQELLCEGDL-----GISFNF 785
Qy 825 DTTVFIESTRORVAEATLENRGENAYSTVLNISQSANLQFASLI---QKEDSDGSIEC 881
Db 786 SGLQVLVVGSGPELTVTVTWNEGEDSYGLTVKFTYPAGLSYRRVTGTQCPHYPLRLAC 845
Qy 882 VNE---ERRLOQVCNVSYPFRRAKAKVAFLRDSFBSKSIPL-HHLEIELAAGSDSNED 937
Db 846 EAEPAQEDLRSSCSINHPIFREGAKTFTMTFVSYKAFGLCDRLLLRAKASSENKPD 905
Qy 938 STKEDNVAFLRPHLYEADVLFRSSSLSHYEVKLNLSLERYDGGIPPPSCIFRIONLGL 997
Db 906 TNK--TAPQLELPVKYTVTVTLISROEDSTNH-VNFSSS---HGGRRQEAHRYRVNLS 959
Qy 998 FPIHGMKMITIPIATRSNRLLKLRLDFTLDEANTSCNIGWNSTEYRTPVEE-----DL 1052
Db 960 LKL-AVRVNFVFEVL-----LNGVAVMDVTLSSPAQGVSCVSOQKPPQNPDLTQI 1009
Qy 1053 RRAPOLNHSNDVSNINCIRLVNQ-EINPFLGLN---WLRSI---KALKYKMKIMV 1105
Db 1010 QRSVLDSCIADCLHRCDFIPSLDQDELDFILRNLNLSFGWSQTLQEKVLLVSAEITF 1069
Qy 1106 NAALQROHSPFIREDPSRQIEFISKOEDQV--PIWITVSTGLGLLLALLVLA 1163
Db 1070 DTSVYSQLPQEAFLR---AQVETTL---EYVVYEFIFLVAGSVGGLLLALLITVVL 1122
Qy 1164 RKLGHFRSARRRREPGLD 1181
Db 1123 YKLGF---XKQYKEMLD 1137

RESULT 13

US-09-350-259-55
; Sequence 55, Application US/09350259
; Patent No. 6620915
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; TITLE OF INVENTION: No. 6620915el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/350,259
; EARLIER FILING DATE: 1999-07-08
; EARLIER APPLICATION NUMBER: 09/193,043
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus

US-09-350-259-55

Query Match

19.0%; Score 1184.5; DB 4; Length 1161;

Best Local Similarity 29.6%; Pred. No. 1.8e-97;
Matches 361; Conservative 207; Mismatches 519; Indels 131; Gaps 43;
Qy 7 LVWAWALSWPGFTDFTNMTRKPRVTPGSTAFPGYTVQOHDISGNKWLWVGABLETNG 66
Db 8 LLCGWVLASCHG---SNLDVEEP-IVPREDAASFGQTVVQF---GGSLRLVVGABLEVA 59
Qy 67 YQKTDGVYKCPVIHGNCTKLNGLSVTLNSVSEKDNMRGLSLATNPKNDSFLACSLWS 126
Db 60 VNQGRGLYDCAAPATGMCQPIVL-RSPLEAV-----NMSLGLSLVTATNNAQLLACGPTAQ 113
Qy 127 HECSSYYTTCMCGRVNSNFRSKTAPALQRC-QTYMDIVIVLDGNSI--YPWVEVQH 183
Db 114 RACVKNMYAKSGCLLGGSLQFIQAVPASMPCEPQEMDIAFLIDGSGINQORDFAQMKD 173
Qy 184 FLINILKFPYTGPGQIQGVVVOYGDVVBHEHLNDYRSVKDVVEAASHIEQGGTETITA 243
Db 174 FVKALMGEP--ASTSTLFSLMQYSNLIKTHFTTTEFKNILDPQSLVDPIVLOQGL-TYTA 230
Qy 244 FGIEPARSEAF--QKGGKRGKAKVMIVITDGEHSDP-DLEKVIQOOSRDNVTRVAVVL 300
Db 231 TGIRTWEEELFHSKNGSRKSAKILLVITDQKYPDPLEYSDVIFAADKAGIIRVIGVG 290
Qy 301 GYINRRGINPETFLNEIKYIASDPDDKHFFNVNTEAALKDVIDALGDRIFSELT-NKNE 359
Db 291 DAFQB-----PTALKELNTIGSAPPQDHVKVGNFAALRSIQRLQOEKIFATEGTQSRSS 345
Qy 360 TSFGLMSQTSFSSHVHVEDGVLGNAGVAYDNGAVLKETSAGKVP--LRSEYLFKEPPEE 417
Db 346 SSFQHEMSQBFSSALTSQDGPVLGAVGVSFGSGGAF-----LYPNTNTRPTFNMSQIN 398
Qy 418 LKNHGAYLGYTVTSVVSRRQGRVYVAGAPRNFHTGKVLFTMNNRSLTIHOAMGQQTIG 477
Db 399 VDMRDSYLGYS-TAVAFWKGVHSLILGAPRHQHTCKWVFT-QEABHWRPKSEVRGTQIG 456
Qy 478 SYFGSEITSDVIDGQVTDVLLVAGPMYFNEGRERGVYVYELR--QNRVYVNGTLKOSH 535
Db 457 SYFGASLCSVDVDRDGSXDLVLIGAPHYVEQTR-GQVSVFVPGVGRVQCEATLHGRO 515
Qy 536 SYQNARFGSSASVREDLNQDSYNDVVGCAPIEDNHAGAIYIFHG-FRGSILKTPQRITA 594
Db 516 GHPWGRFGVALTVLGDVNGDNLADVAIGAPEEESRGAVYIFHGASRLIEIMPSPQRVTG 575
Qy 595 SELATGLQVFGCSIHGQDLNEDGLDLAVAGALGNVILWSPRVVQINASHFEPSKINI 654
Db 576 SLSRLQYFGQSLGGQDLTQDGLVDLAVGAQGHVLLLSRLPLLLKVELSIRFAPMEYAK 635
Qy 655 FHRDCKRGRDATCL---AFLCFTPIFLAPHFQTTTVG-IRYNATMDERRYTTPRAHLDE 710
Db 636 AVYQCWE--RTPTVLEAGEATVCLTVHKGSPDLLGNVQGSVRYDLALDPGRLLISRAIFDE 693
Qy 711 GGDRTNRAVLSSSQELCERINPHVLD-TADYVKPVTFSVEYSI-----EDPDHGMULD 764
Db 694 TKNCTLTGRKTLGLGDH-CETVKLLLPDCVEDAVSPFIILRLNFSLVRSASPRNLUHPVLA 752
Qy 765 DGMPTTLRVSPFMWGNCDNEHCVPDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTLSF 824
Db 753 VGSQDHIATSLPFKNCKQELLCEGDL-----GISFNF 785
Qy 825 DTTVFIESTRORVAEATLENRGENAYSTVLNISQSANLQFASLI---QKEDSDGSIEC 881
Db 786 SGLQVLVVGSGPELTVTVTWNEGEDSYGLTVKFTYPAGLSYRRVTGTQCPHYPLRLAC 845
Qy 882 VNE---ERRLOQVCNVSYPFRRAKAKVAFLRDSFBSKSIPL-HHLEIELAAGSDSNED 937
Db 846 EAEPAQEDLRSSCSINHPIFREGAKTFTMTFVSYKAFGLCDRLLLRAKASSENKPD 905
Qy 938 STKEDNVAFLRPHLYEADVLFRSSSLSHYEVKLNLSLERYDGGIPPPSCIFRIONLGL 997
Db 906 TNK--TAPQLELPVKYTVTVTLISROEDSTNH-VNFSSS---HGGRRQEAHRYRVNLS 959
Qy 998 FPIHGMKMITIPIATRSNRLLKLRLDFTLDEANTSCNIGWNSTEYRTPVEE-----DL 1052

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:52:17 ; Search time 1.15642 Seconds
(without alignments)
1913.143 Million cell updates/sec

Title: US-09-980-403-2_COPY_804_826

Perfect score: 125

Sequence: 1 EYQVRLKPAQDCSAYTSLSFD 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	41.6	198	2	T25436
2	49	39.2	148	2	C82337
3	48	38.4	1184	2	T09484
4	47	37.6	335	2	T37262
5	47	37.6	398	2	A35281
6	47	37.6	612	2	T39684
7	47	37.6	771	2	T26176
8	46	36.8	223	2	B71498
9	46	36.8	235	2	F81558
10	46	36.8	330	2	D82756
11	46	36.8	572	2	T32303
12	46	36.8	782	2	T90742
13	46	36.8	782	2	C85593
14	46	36.8	782	2	A64821
15	46	36.8	1275	2	T41523
16	45.5	36.4	575	2	B69450
17	45	36.0	566	1	B41648
18	45	36.0	574	2	I66868
19	45	36.0	576	2	A42336
20	45	36.0	785	2	T11719
21	44	35.2	645	2	S51880
22	43.5	34.8	365	2	T06991
23	43.5	34.8	365	2	T06990
24	43.5	34.8	1312	2	D85066
25	43	34.4	328	2	H84548
26	43	34.4	345	2	A87338
27	43	34.4	373	2	AG3334
28	43	34.4	480	2	T46047
29	43	34.4	637	2	T37713

30 43 34.4 683 2 T10720
31 43 34.4 924 1 S13913
32 43 34.4 1276 2 T09204
33 43 34.4 1291 2 T09273
34 43 34.4 1441 2 A88355
35 43 34.4 1737 2 T19606
36 43 34.4 2910 2 T42214
37 42 33.6 200 2 G82270
38 42 33.6 257 2 T03136
39 42 33.6 493 2 T06031
40 42 33.6 792 2 S72831
41 42 33.6 1054 2 T30177
42 42 33.6 1716 2 T14103
43 42 33.6 1717 2 T13961
44 42 33.6 2491 1 A28372
45 41.5 33.2 521 2 T18896

ALIGNMENTS

RESULT 1

T25436

hypothetical protein T28H10.2 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T25436

R/Kershaw, J.

submitted to the EMBL Data Library, June 1996

A/Reference number: Z20034

A/Accession: T25436

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-198 <WIL>

A/Cross-references: EMBL:Z75551; PIDN:CAA9934.1; GSPDB:GN00023; CESP:T28H10.2

A/Experimental source: clone T28H10

C/Genetics:

A/Gene: CESP:T28H10.2

A/Map position: 5

A/Introns: 24/1; 126/2

Query Match 41.6%; Score 52; DB 2; Length 198;

Best Local Similarity 43.5%; Pred. No. 0.94; Mismatches 3; Indels 10; Gaps 0;

Matches 10; Conservative 3; Mismatches 10; Indels 10; Gaps 0;

Qy 1 EYQVRLKPAQDCSAYTSLSFD 23

Db 49 EYIKNCRKSCGNCPRYLKFDI 71

RESULT 2

C82337

hypothetical protein VC0317 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C:Species: *Vibrio cholerae*

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C/Accession: C82337

R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A/Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A/Reference number: A82035; MUID:20406933; PMID:10952301

A/Accession: C82337

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-148 <HEI>

A/Cross-references: GB:AE004120; GB:AE003852; NID:9654727; PIDN:AAF93490.1; GSPDB:GN001

A/Experimental source: serogroup O1; strain N16961; biotype El Tor

C/Genetics:

A/Gene: VC0317

A/Map position: 1

Query Match 39.2%; Score 49; DB 2; Length 148;

Best Local Similarity 50.0%; Pred. No. 2.2; Indels 7; Mismatches 2; Gaps 0;

Matches 9; Conservative 2;

Qy 1 EYQQRVLRKPAQDCSAYT 18

Db 86 DYCCQKQWLNPAFCFAYT 103

RESULT 3

T09484

cartilage intermediate layer protein precursor - human

C:Species: Homo sapiens (man)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T09484

R:Lorenzo, P.; Neame, P.; Sommarin, Y.; Heinegard, D.

J. Biol. Chem. 273, 23469-23475, 1998

A:Title: Cloning and deduced amino acid sequence of a novel cartilage protein (CILP) id

A:Reference number: Z16689; MUID:98389785; PMID:9722584

A:Accession: T09484

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1184 <LOR>

A:Cross-references: EMBL:AF035408; NID:G3513502; PIDN:AAC33838.1; PID:G3513503

A:Experimental source: tissue type articular cartilage

C:Genetics:

A>Note: CILP

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-1184/Product: cartilage intermediate layer protein #status predicted <MAT>

Query Match 38.4%; Score 48; DB 2; Length 1184;

Best Local Similarity 40.0%; Pred. No. 22;

Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 2 YCQVRVLRKPAQDCSAYTISF 21

Db 118 WCLNREQRPGQNCNYSYTVRF 137

RESULT 4

T37262

probable tyrosine kinase receptor W04G5.6A - Caenorhabditis elegans (fragment)

C:Species: Caenorhabditis elegans

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T37262

R:Popovici, C.; Roubin, R.; Coulter, F.; Pontarotti, P.; Birnbaum, D.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z21654

A:Accession: T37262

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-335 <POP>

A:Cross-references: EMBL:AF188749; PIDN:AAF00546.1

A:Experimental source: strain N2

C:Genetics:

A:Map position: 1

A>Note: W04G5.6A

Query Match 37.6%; Score 47; DB 2; Length 335;

Best Local Similarity 38.1%; Pred. No. 9.7;

Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 2 YCQVRVLRKPAQDCSAYTISFD 22

Db 38 YRAKLRRKPTQVSSHSIYFD 58

RESULT 5

A35281

integumentary mucin B.1 - African clawed frog (fragment)

C:Species: Xenopus laevis (African clawed frog)

C>Date: 17-Aug-1990 #sequence_revision 06-Nov-1992 #text_change 03-Nov-2000

C:Accession: A35281

R:Probst, J.C.; Gertzen, E.M.; Hoffmann, W.

Biochemistry 29, 6240-6244, 1990

A:Title: An integumentary mucin (FIM-B.1) from Xenopus laevis homologous with von Wille

A:Reference number: A35281; MUID:91002513; PMID:2207068

A:Accession: A35281

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-398 <PRO>

A:Cross-references: GB:J02910; NID:G214145; PIDN:AAA49711.1; PID:G214146

C:Superfamily: pig submaxillary mucin

Query Match 37.6%; Score 47; DB 2; Length 398;

Best Local Similarity 57.1%; Pred. No. 11;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 4 QRVLRKPAQDCSAY 17

Db 219 RRVLRKPKKCCGY 232

RESULT 6

T39684

hypothetical protein SPBC1778.02 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39684

R:Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, March 1998

A:Reference number: Z21869

A:Accession: T39684

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-612 <OLI>

A:Cross-references: EMBL:AL049489; PIDN:CAB39797.1; GSPDB:GN00067; SPDB:SPBC1778.02

A:Experimental source: strain 972h-; cosmid cl778

C:Genetics:

A:Gene: SPDB:SPEC1778.02

A:Map position: 2

A:Introns: 75/3

Query Match 37.6%; Score 47; DB 2; Length 612;

Best Local Similarity 47.4%; Pred. No. 17;

Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 2 YCQVRVLRKPAQDCSAYTIS 20

Db 106 YCQRIIVAPYSSQKDYTQS 124

RESULT 7

T26176

hypothetical protein W04G5.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26176

R:Kershaw, J.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z20164

A:Accession: T26176

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-771 <WIL>

A:Cross-references: EMBL:Z93391; PIDN:CAB07686.1; GSPDB:GN00019; CESP:W04G5.6

A:Experimental source: clone W04G5

C:Genetics:

A:Gene: CESP:W04G5.6

A:Map position: 1

A:Introns: 13/1; 116/1; 157/3; 208/3; 245/2; 384/3; 395/1; 438/2; 530/1; 570/2; 627/3;

Query Match 37.6%; Score 47; DB 2; Length 771;

Best Local Similarity 38.1%; Pred. No. 21;

Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 2 YCQVRVLRKPAQDCSAYTISFD 22

RESULT 10
D82756
pyridoxal phosphate biosynthetic protein XF0839 [imported] - Xylella fastidiosa (strain
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: D82756
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 451-457, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; PMID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: D82756

hypothetical protein ECs0312 [imported] - *Escherichia coli* (strain O157:H7, substrain R1)
H90742
C:Species: *Escherichia coli*
C:date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: H90742
R:Hayashi, N.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuahara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:44:02 ; Search time 0.674581 Seconds
(without alignments)
1775.344 Million cell updates/sec

Title: US-09-980-403-2_COPY_804_826
Perfect score: 125
Sequence: 1 EYQVLRKPAQDCSAVTLSPDT 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	125	100.0	1189	1 ITAH HUMAN	Q9ux5 homo sapien
2	51	40.8	7059	1 RIAB CVBQ	Q8v6w7 b replicase
3	51	40.8	7094	1 RIAB CVBEN	Q8v6w7 b replicase
4	51	40.8	7094	1 RIAB CVBLU	Q8v6w7 b replicase
5	51	40.8	7094	1 RIAB CVBM	Q8v6w7 b replicase
6	47	37.6	398	1 MBL XENLA	P38565 xenopus lae
7	47	37.6	693	1 RAP1 SCHPO	Q96tl7 schizosacch
8	46	36.8	330	1 PDXA XYLFA	Q87aj1 xylella fas
9	46	36.8	330	1 PDXA XYLFT	Q87aj1 xylella fas
10	46	36.8	782	1 YLIE ECOLI	P75800 escherichia
11	45.5	36.4	575	1 SYP ARCTU	Q28664 archaeoglob
12	45	36.0	574	1 MP12 RAT	P48966 rattus norv
13	45	36.0	576	1 MP12 MOUSE	P30306 mus musculu
14	45	36.0	580	1 MP12 HUMAN	P30305 homo sapien
15	45	36.0	785	1 VP35 SCHPO	Q74552 schizosacch
16	45	36.0	1093	1 SM5B MOUSE	Q60519 mus musculu
17	44	35.2	841	1 HS74 MOUSE	Q61316 mus musculu
18	43	34.4	637	1 YDOD SCHPO	Q3733 schizosacch
19	43	34.4	924	1 HXK3 RAT	P37926 rattus norv
20	42.5	34.0	229	1 NEP1 CANGA	Q96cp2 candida gla
21	42.5	34.0	651	1 ABG5 HUMAN	Q9h222 homo sapien
22	42	33.6	317	1 CA5B HUMAN	Q9v2d0 homo sapien
23	42	33.6	493	1 HXKL ARATH	Q9t071 arabidopsis
24	42	33.6	792	1 YV01 MYCLE	Q49736 mycobacteri
25	42	33.6	1716	1 REA1 RAT	Q54889 rattus norv
26	42	33.6	1717	1 REA1 HUMAN	Q95602 homo sapien
27	42	33.6	1717	1 REA1 MOUSE	Q35134 mus musculu
28	42	33.6	2491	1 MPRI HUMAN	P11717 homo sapien
29	41	32.8	401	1 KICS POTR	P21856 potocous tr
30	41	32.8	419	1 ASSY DROME	Q97069 drosophila
31	41	32.8	1110	1 Y256 HUMAN	Q93073 homo sapien
32	41	32.8	2186	1 YL52 CAEBL	P34431 caenorhabdi
33	41	32.8	7073	1 RIAB CVHSA	P59641 h replicase

34	40.5	32.4	267	1 NEP1 CANAL	Q9p8p7 candida alb
35	40.5	32.4	595	1 PRIM CHLTR	O84799 chlamydia t
36	40.5	32.4	652	1 GP63 CRIFA	Q06031 crithidia f
37	40	32.0	185	1 BY55 MOUSE	O88875 mus musculu
38	40	32.0	295	1 SM4D CHICK	Q90665 gallus gall
39	40	32.0	336	1 TTC9 HUMAN	Q92623 homo sapien
40	40	32.0	623	1 DNK2 CYAPA	Q37106 cyanophora
41	40	32.0	634	1 DNK2 SYNPF	P50021 synechococc
42	40	32.0	690	1 NCPR PHAAU	P37116 phaseolus a
43	40	32.0	696	1 EFG RHIL0	Q9en59 thizobium i
44	40	32.0	989	1 IDE DROME	P22817 drosophila
45	40	32.0	1293	1 PUR4 SHEON	Q8ec57 shewanella

ALIGNMENTS

RESULT 1

ID	ITAH HUMAN	STANDARD	PRT	1189 AA.
AC	Q9UX5; Q9UKQ1;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Integrin alpha-11 precursor.			
GN	ITGA11.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal heart, and Osteoblast;			
RA	MEDLINE=99417678; PubMed=10486209;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,			
RA	Wang S.-X., Morris C.M., Krissansen G.W.;			
RT	"Cloning, sequence analysis, and chromosomal localization of the novel			
RT	human integrin alphall subunit (ITGA11).";			
RL	Genomics 60:179-187(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal muscle, and Uterus;			
RA	MEDLINE=99395147; PubMed=10464311;			
RA	Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;			
RT	"CDNA Cloning and Chromosomal Localization of Human alpha(11)			
RT	Integrin. A collagen-binding, i domain-containing, beta(1)-associated			
RT	integrin alpha-chain present in muscle tissues.";			
RL	J. Biol. Chem. 274:25735-25742(1999).			
RN	[3]			
RP	SEQUENCE OF 954-1188 FROM N.A.			
RC	TISSUE=Fibroblast;			
RA	Andreu N., Estivill X., Escarceller M., Sunoy L.;			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.			
CC	-!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11			
CC	ASSOCIATES WITH BETA-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND			
CC	HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO			
CC	LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO			
CC	FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS			
CC	SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN			
CC	PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.			
CC	-!- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING			
CC	FETAL MUSCLE CELLS (IN VITRO).			
CC	-!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS			
CC	WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 1 VWFA domain.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF109681; AAF01258.1; -
EMBL; AF137378; AAD51919.2; -
EMBL; AL359064; CAB94392.1; -
HSSP; P17301; IAOX.
DR Genew; HGNC:6136; ITGAl1.
MIM; 604789; -
GO; GO:0008305; C:integrin complex; TAS.
GO; GO:0004895; F:cell adhesion receptor activity; TAS.
GO; GO:0005518; F:collagen binding; TAS.
GO; GO:0007160; P:cell-matrix adhesion; TAS.
GO; GO:0007517; P:muscle development; TAS.
InterPro; IPR000413; Integrin_alpha.
InterPro; IPR002035; VWF_A.
Pfam; PF01839; FG-GAP; 3.
Pfam; PF00092; vwa; 1.
PRINTS; PRO1185; INTEGRINA.
PRINTS; PRO0453; VWFADOMAIN.
SMART; SM00191; Int_alpha; 5.
SMART; SM00327; VWA; 1.
PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
PROSITE; PS00234; VWFA; 1.
Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
Signal; Repeat; Calcium; Magnesium; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 1189
FT DOMAIN 23 1142
FT TRANSMEM 1143 1165
FT DOMAIN 1166 1189
FT REPEAT 38 94
FT REPEAT 102 163
FT DOMAIN 167 345
FT REPEAT 359 420
FT REPEAT 422 475
FT REPEAT 477 537
FT REPEAT 539 598
FT REPEAT 601 653
FT DOMAIN 1154 1162
FT DOMAIN 1174 1177
FT CA_BIND 488 496
FT CA_BIND 551 559
FT CA_BIND 613 621
FT DISULFID 76 83
FT DISULFID 121 139
FT DISULFID 129 159
FT DISULFID 659 688
FT DISULFID 674 729
FT DISULFID 781 787
FT DISULFID 881 893
FT CARBOHYD 82 95
FT CARBOHYD 95 95
FT CARBOHYD 291 291
FT CARBOHYD 331 331
FT CARBOHYD 338 338
FT CARBOHYD 449 449
FT CARBOHYD 462 462
FT CARBOHYD 528 528
FT CARBOHYD 642 642
FT CARBOHYD 694 694
FT CARBOHYD 857 857
FT CARBOHYD 894 894
FT CARBOHYD 973 973
FT CARBOHYD 1032 1032
FT CARBOHYD 1040 1040
FT VARIANT 433 433
FT VARIANT 524 524
FT VARIANT 524 524
R -> L.
/FTId=VAR_009889.
/FTId=VAR_009890.

FT VARIANT 972 972 L -> P.
FT VARIANT 1003 1003 /FTId=VAR_009891.
FT VARIANT 1030 1030 I -> M.
FT VARIANT 1030 1030 /FTId=VAR_009892.
FT VARIANT 1094 1094 Missing.
FT VARIANT 1094 1094 /FTId=VAR_009893.
FT VARIANT 1094 1094 L -> V.
FT VARIANT 1094 1094 /FTId=VAR_009894.
SQ SEQUENCE 1189 AA; 133609 MW; 60303C08AA4CD52 CRC64;
Query Match 100.0%; Score 125; DB 1; Length 1189;
Best Local Similarity 100.0%; Pred. No. 9.4e-12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EYQQRVLKPKAQCDSAYTLSPDT 23
DB 804 EYQQRVLKPKAQCDSAYTLSPDT 826
RESULT 2
ID RIAB CVBQ STANDARD; PRT; 7059 AA.
AC Q8V6W7; Q8V6W6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Replicase polypeptide lab (pp1ab) (ORF1a) [Includes:
DE Replicase polypeptide lab (pp1ab) (ORF1a) [Contains: p28; p65; p210
DE (EC 3.4.24.-) (Papain-like proteinases 1/2) (PL1-PRO) (PL2-PRO); Peptide
DE HD2 (p44); 3C-like proteinase (EC 3.4.24.-) (3CL-PRO) (3CLP) (M-PRO)
DE (p27); Unknown protein 1; p10; p22; p12; Growth factor-like peptide
DE (GFL) (p15); RNA-directed RNA polymerase (EC 2.7.7.48) (RdRp) (Pol)
DE (p100); Helicase (Hel) (p67); Unknown protein 2; p35; Unknown protein
DE 3].
OS Bovine coronavirus (strain Quebec) (BCoV) (BCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxID=11133;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoo D., Pei Y., Parker M.D., Cox G.J.;
RA "Bovine coronavirus (Quebec strain) full-length genomic sequence.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: The replicase polypeptide of coronaviruses is a
CC multifunctional protein: it contains the activities necessary for
CC the transcription of negative stranded RNA, leader RNA, subgenomic
CC mRNAs and progeny virion RNA as well as proteinases responsible
CC for the cleavage of the polypeptide into functional products.
CC -!- FUNCTION: The papain-like proteinase 1 (PL1-PRO) and papain-like
CC proteinase 2 (PL2-PRO) are responsible for the cleavages located
CC at the N-terminus of the replicase polypeptide. Activity of PL1-
CC PRO is strongly dependent on Zn(2+) (By similarity).
CC -!- FUNCTION: The main proteinase 3CL-PRO is responsible for the
CC majority of cleavages as it cleaves the C-terminus of replicase
CC polypeptide at 11 sites. Recognizes substrates containing the core
CC sequence [ILMV]-Q-[SAGC]. Inhibited by the substrate-analog
CC Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK (By similarity).
CC -!- FUNCTION: The helicase which contains a zinc finger structure
CC displays RNA and DNA duplex-unwinding activities with 5' to 3'
CC polarity. ATPase activity is strongly stimulated by poly(U),
CC poly(GT), poly(C), poly(dA), but not by poly(G) (By similarity).
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: 3CL-PRO exists as monomer and homodimer. The Cys-rich
CC growth factor-like peptide forms a homodimer; disulfide-linked (By
CC similarity).
CC -!- DOMAIN: The hydrophobic domains (HD) could mediate the membrane
CC association of the replication complex and thereby alter the
CC architecture of the host cell membrane.
CC -!- PTM: Specific enzymatic cleavages in vivo by its own proteases
CC yield mature proteins. 3CL-PRO and PL-PRO proteinases are
CC autocatalytically processed (By similarity).
CC -!- MISCELLANEOUS: This protein is translated as a 1A-1B polypeptide

RESULT 4
 RIAB_CVBLU STANDARD; PRT; 7094 AA.
 ID QV439; QV440;
 AC 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Replicase polypeptide 1a (pp1a) (ORF1ab) [ORF1ab polypeptide] [Includes:
 DE Replicase polypeptide 1a (pp1a) (ORF1a) [Contains: p28; p65; p210
 DE (EC 3.4.24.-) (Papain-like proteinases 1/2) (PL1-PRO/PL2-PRO); Peptide
 DE HD2 (p44); 3C-like proteinase (EC 3.4.24.-) (3CL-PRO) (3CLP) (M-PRO)
 DE (p27); Unknown protein 1; p10; p22; p12; Growth factor-like peptide
 DE (GFL) (p15); RNA-directed RNA polymerase (EC 2.7.7.48) (RdRp) (Pol)
 DE (p100); Helicase (Hel) (p67); Unknown protein 2; p35; Unknown protein
 DE 3].
 OS Bovine coronavirus (strain 98TXSF-110-LUN) (BCoV-LUN) (BCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=233264;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21571638; PubMed=11714968;
 RA Chouljenko V.N., Lin X.Q., Storz J., Kousoulas K.G., Gorbalenya A.E.;
 RT "Comparison of genomic and predicted amino acid sequences of
 RT respiratory and enteric bovine coronaviruses isolated from the same
 RT animal with fatal shipping pneumonia.";
 RL J. Gen. Virol. 82:2927-2933(2001).
 CC -!- FUNCTION: The replicase polypeptide of coronaviruses is a
 CC multifunctional protein: it contains the activities necessary for
 CC the transcription of negative stranded RNA, leader RNA, subgenomic
 CC mRNAs and progeny virion RNA as well as proteinases responsible
 CC for the cleavage of the polypeptide into functional products.
 CC -!- FUNCTION: The papain-like proteinase 1 (PL1-PRO) and papain-like
 CC proteinase 2 (PL2-PRO) are responsible for the cleavages located
 CC at the N-terminus of the replicase polypeptide. Activity of PL1-
 CC PRO is strongly dependent on Zn(2+) (By similarity).
 CC -!- FUNCTION: The main proteinase 3CL-PRO is responsible for the
 CC majority of cleavages as it cleaves the C-terminus of replicase
 CC polypeptide at 11 sites. Recognized substrates containing the core
 CC sequence [ILMV]-Q-[SAGC]. Inhibited by the substrate-analog
 CC Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK (By similarity).
 CC -!- FUNCTION: The helicase which contains a zinc finger structure
 CC displays RNA and DNA duplex-unwinding activities with 5' to 3'
 CC polarity. ATPase activity is strongly stimulated by poly(U),
 CC poly(dT), poly(C), poly(GA), but not by poly(G) (By similarity).
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -!- SUBUNIT: 3CL-PRO exists as monomer and homodimer. The Cys-rich
 CC growth factor-like peptide forms a homodimer; disulfide-linked (By
 CC similarity).
 CC -!- DOMAIN: The hydrophobic domains (HD) could mediate the membrane
 CC association of the replication complex and thereby alter the
 CC architecture of the host cell membrane.
 CC -!- PM: Specific enzymatic cleavages in vivo by its own proteases
 CC yield mature proteins. 3CL-PRO and PL-PRO proteinases are
 CC autocatalytically processed (By similarity).
 CC -!- MISCELLANEOUS: This protein is translated as a 1a-1b polypeptide
 CC by a ribosomal frameshifting mechanism. The 1a polypeptide is
 CC produced by itself but not the 1b polypeptide.
 CC -!- SIMILARITY: Contains 2 peptidase family C16 domains.
 CC -!- SIMILARITY: Contains 1 peptidase family C30 domain.
 CC -!- SIMILARITY: Contains 1 Alpp domain.
 CC -!- SIMILARITY: Contains 2 C4-type zinc fingers.
 CC -!- SIMILARITY: Contains 1 viral helicase-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC CC
 DR EMBL; AF391542; AAL57305.1; ALT_SEQ.
 DR EMBL; AF391542; AAL57315.1; -.
 DR MEROPS; C16.001; -.
 DR MEROPS; C16.006; -.
 DR MEROPS; C30.001; -.
 DR InterPro; IPR002589; Alpp.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR002705; Peptidase_C16.
 DR InterPro; IPR008740; Peptidase_C30.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSTir.
 DR InterPro; IPR006060; Viral_helicase.
 DR Pfam; PF01861; Alpp; 1.
 DR Pfam; PF01831; Peptidase_C16; 1.
 DR Pfam; PF05409; Peptidase_C30; 1.
 DR Pfam; PF01443; Viral_helicase; 1.
 DR SMART; SM00506; Alpp; 1.
 DR Hydrolase; Transferase; Helicase; RNA-directed RNA polymerase;
 DR Thiol protease; ATP-binding; Metal-binding; Zinc; Repeat;
 KW Transmembrane; Zinc-finger; Polypeptide; Ribosomal frameshift.
 KW CHAIN 1 246
 FT CHAIN 247 851
 FT CHAIN 852 2750
 FT CHAIN 2751 3246
 FT CHAIN 3247 3549
 FT CHAIN 3550 3836
 FT CHAIN 3837 3925
 FT CHAIN 3926 4122
 FT CHAIN 4123 4232
 FT CHAIN 4233 4369
 FT CHAIN 4370 5297
 FT CHAIN 5298 5900
 FT CHAIN 5901 6421
 FT CHAIN 6422 6795
 FT CHAIN 6796 7094
 FT CHAIN 7138 2158
 FT TRANSMEM 2199 2219
 FT TRANSMEM 2227 2247
 FT TRANSMEM 2313 2333
 FT TRANSMEM 2343 2363
 FT TRANSMEM 2365 2385
 FT TRANSMEM 2752 2772
 FT TRANSMEM 2824 2844
 FT TRANSMEM 3009 3029
 FT TRANSMEM 3031 3051
 FT TRANSMEM 3063 3083
 FT TRANSMEM 3090 3110
 FT TRANSMEM 3115 3135
 FT TRANSMEM 3558 3578
 FT TRANSMEM 3588 3608
 FT TRANSMEM 3614 3634
 FT TRANSMEM 3657 3677
 FT TRANSMEM 3684 3704
 FT TRANSMEM 3711 3731
 FT TRANSMEM 3755 3775
 FT TRANSMEM 1055 1257
 FT DOMAIN 1297 1407
 FT DOMAIN 1652 1863
 FT DOMAIN 2138 2385
 FT DOMAIN 2752 3135
 FT DOMAIN 3558 3775
 FT DOMAIN 4908 5210
 FT DOMAIN 5574 5872
 FT ZN_FING 1151 1179
 FT ZN_FING 1749 1785
 FT ZN_FING 5302 5372
 FT NP_BIND 5578 5585
 FT SITE 246 247
 FT SITE 851 852

FT SITE 2750 2751 CLEAVAGE (BY PL2-PRO) (BY SIMILARITY).
 FT SITE 3246 3247 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 3549 3550 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 3836 3837 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 3925 3926 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 4122 4123 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 4232 4233 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 4369 4370 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 5293 5298 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 5300 5301 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 6421 6422 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT SITE 6795 6796 CLEAVAGE (BY 3CL-PRO) (BY SIMILARITY).
 FT ACT_SITE 1074 1074 PL1-PRO (BY SIMILARITY).
 FT ACT_SITE 1225 1225 PL1-PRO (BY SIMILARITY).
 FT ACT_SITE 1671 1671 PL2-PRO (BY SIMILARITY).
 FT ACT_SITE 1828 1828 PL2-PRO (BY SIMILARITY).
 FT ACT_SITE 3287 3287 3CL-PRO (BY SIMILARITY).
 FT ACT_SITE 3391 3391 3CL-PRO (BY SIMILARITY).
 FT VARIANT 4378 4383 RVGTGS -> GFGVRV (in ORF1a).
 FT VARIANT 4384 7094 MISSING (in ORF1a).
 SQ SEQUENCE 7094 AA; 797335 MW; 00AA5C81C94327C8 CRC64;

Query Match 40.8%; Score 51; DB 1; Length 7094;
 Best Local Similarity 52.6%; Pred. No. 29;
 Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 QRVLRKPQDCSAYTILSPD 22
 |||: |||: |||: |||: |||:
 Db 6212 QRVMLKAAMLCNRYTLCYD 6230

RESULT 5
 RIAB_CVEM STANDARD; PRT: 7094 AA.
 AC Q66198; Q9WQ81;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Replicase polypeptide lab (pplab) (ORF1ab polypeptide) [Includes:
 DE Replicase polypeptide 1a (ppl1) (ORF1a) [Contains: p28; p65; p210
 DE (EC 3.4.24.-) (Papain-like proteinases 1/2) (PL1-PRO/PL2-PRO); Peptide
 DE HD2 (p44); 3C-like proteinase (EC 3.4.24.-) (3CL-PRO) (3CLp) (M-PRO)
 DE (p27); Unknown protein 1; p10; p22; p12; Growth factor-like peptide
 DE (GFL) (p15); RNA-directed RNA polymerase (EC 2.7.7.48) (RdRp) (Pol)
 DE (p100); Helicase (Hel) (p67); Unknown protein 2; p35; Unknown protein
 DE 3].
 OS Bovine coronavirus (strain Mebus) (BCoV) (BCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=11132;
 RN [1]_TaxID=11132;
 RP SEQUENCE FROM N.A.
 RA Brian D.A.;
 RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP REVISIONS.
 RA Brian D.A.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RX SEQUENCE OF 4871-5177 FROM N.A.
 RA MEDLINE=99319897; PubMed=10392726;
 RA Stephenson C.B., Casebolt D.B., Gangopadhyay N.N.;
 RT "Phylogenetic analysis of a highly conserved region of the polymerase
 RT gene from 11 coronaviruses and development of a consensus polymerase
 RT chain reaction assay.";
 RL Virus Res. 60:181-189(1999).
 CC -!- FUNCTION: The replicase polypeptide of coronaviruses is a
 CC multifunctional protein: it contains the activities necessary for
 CC the transcription of negative stranded RNA, leader RNA, subgenomic
 CC mRNAs and progeny virion RNA as well as proteinases responsible
 CC for the cleavage of the polypeptide into functional products.
 CC -!- FUNCTION: The papain-like proteinase 1 (PL1-PRO) and papain-like
 CC proteinase 2 (PL2-PRO) are responsible for the cleavages located

at the N-terminus of the replicase polypeptide. Activity of PL1-
 PRO is strongly dependent on Zn(2+) (By similarity).
 -!- FUNCTION: The main proteinase 3CL-PRO is responsible for the
 majority of cleavages as it cleaves the C-terminus of replicase
 polypeptide at 11 sites. Recognizes substrates containing the core
 sequence [ILMV]-Q-[SAGC]. Inhibited by the substrate-analog
 Chz-Val-Asn-Ser-Thr-Leu-Gln-CMK (By similarity).
 -!- FUNCTION: The helicase which contains a zinc finger structure
 displays RNA and DNA duplex-unwinding activities with 5' to 3',
 polarity. Aprase activity is strongly stimulated by poly(U),
 poly(DT), poly(C), poly(GA), but not by poly(G) (By similarity).
 -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA} (N).
 -!- SUBUNIT: 3CL-PRO exists as monomer and homodimer. The Cys-rich
 growth factor-like peptide forms a homodimer; disulfide-linked (By
 similarity).
 -!- DOMAIN: The hydrophobic domains (HD) could mediate the membrane
 association of the replication complex and thereby alter the
 architecture of the host cell membrane.
 -!- PTM: Specific enzymatic cleavages in vivo by its own proteases
 yield mature proteins. 3CL-PRO and PL-PRO proteinases are
 autocatalytically processed (By similarity).
 -!- MISCELLANEOUS: This protein is translated as a 1A-1B polypeptide
 by a ribosomal frameshifting mechanism. The 1A polypeptide is
 produced by itself but not the 1B polypeptide.
 -!- SIMILARITY: Contains 2 peptidase family C16 domains.
 -!- SIMILARITY: Contains 1 peptidase family C30 domain.
 -!- SIMILARITY: Contains 1 Alpp domain.
 -!- SIMILARITY: Contains 2 C4-type zinc fingers.
 -!- SIMILARITY: Contains 1 viral helicase-type zinc finger.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL; U00735; AAA64744.2; -.
 CC EMBL; AF124985; AAD32989.1; -.
 DR MEROPS; C16.001; -.
 DR MEROPS; C16.006; -.
 DR MEROPS; C30.001; -.
 DR InterPro; IPR002589; Alpp.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR007095; RNA_pol_DS_Ps.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01661; Alpp; 1.
 DR Pfam; PF01831; Peptidase_C16; 1.
 DR Pfam; PF05409; Peptidase_C30; 1.
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR SMART; SM00506; Alpp; 1.
 KW Hydrolase; Transferase; Helicase; RNA-directed RNA polymerase;
 KW Thiol protease; ATP-binding; Metal-binding; Zinc; Repeat;
 KW Transmembrane; Zinc-finger; Polypeptide; Ribosomal frameshift.
 FT CHAIN 1 246
 FT CHAIN 247 851
 FT CHAIN 852 2750
 FT CHAIN 2751 3246 PEPTIDE HD2 (BY SIMILARITY).
 FT CHAIN 3247 3549 3C-LIKE PROTEINASE (BY SIMILARITY).
 FT CHAIN 3550 3836 UNKNOWN PROTEIN 1 (BY SIMILARITY).
 FT CHAIN 3837 3925 P10 (BY SIMILARITY).
 FT CHAIN 3926 4122 P22 (BY SIMILARITY).
 FT CHAIN 4123 4232 P12 (BY SIMILARITY).
 FT CHAIN 4233 4369 GROWTH FACTOR-LIKE PEPTIDE (BY
 FT CHAIN 4370 5297 SIMILARITY).
 FT CHAIN 5298 5900 RNA-DIRECTED RNA POLYMERASE (BY
 FT CHAIN 5901 6421 SIMILARITY).
 FT CHAIN 6422 6795 UNKNOWN PROTEIN 2 (BY SIMILARITY).
 FT CHAIN 6796 UNKNOWN PROTEIN 3 (BY SIMILARITY).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 CC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91002513; PubMed=2207068;
 RA Probst J.C., Gertzen E.-M., Hoffmann W.;
 RT "An integumentary mucin (FIM-B.1) from *Xenopus laevis* homologous with
 an Willebrand factor";
 RL Biochemistry 29:6240-6244(1990).
 CC -!- FUNCTION: Could be involved in defense against microbial
 CC infections. Protects the epithelia from external environment.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed and stored exclusively in mature
 CC mucous glands of the skin.
 CC -!- PTM: Extensively O-glycosylated.
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J02910; AAA49711.1; --
 DR F1R; A35281; A35281.
 DR InterPro; IPR006208; Cys_knot.
 DR InterPro; IPR006207; Cys_knot_C.
 DR InterPro; IPR001007; VWFC.
 DR Pfam; PF00007; Cys_knot; 1.
 DR SMART; SM00041; CT; 1.
 DR SMART; SM00214; WVC; 2.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS01184; VWFC_2; 1.
 KW Repeat; Glycoprotein.
 FT NON_TER 1 1
 FT DOMAIN 1 1 93
 FT
 FT 8 X 11 AA TANDEM REPEATS OF G-E-S-T-P-
 FT A-P-S-E-T-T.
 FT 1 (INCOMPLETE).
 FT REPEAT <1 7
 FT REPEAT 8 18
 FT REPEAT 19 29
 FT REPEAT 30 40
 FT REPEAT 41 51
 FT REPEAT 52 62
 FT REPEAT 66 76
 FT REPEAT 83 93
 FT REPEAT 169 234
 FT DOMAIN 301 392
 FT CTCK.
 FT DISULFID 301 356
 FT BY SIMILARITY.
 FT DISULFID 323 370
 FT BY SIMILARITY.
 FT DISULFID 332 386
 FT BY SIMILARITY.
 FT DISULFID 336 388
 FT BY SIMILARITY.
 FT DISULFID ? 391
 FT BY SIMILARITY.
 SQ SEQUENCE 398 AA; 42101 MW; C06C86A805A3C3A4 CRC64;
 Query Match 37.6%; Score 47; DB 1; Length 398;
 Best Local Similarity 57.1%; Pred. No. 6.2;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 4 QRVLRKPAQDCSAY 17
 Db 219 RRVLRKPGKSCGY 232
 RESULT 7
 RAPI SCHPO STANDARD; PRT; 693 AA.
 ID MUB1_XENLA
 AC P38565; STANDARD; PRT; 398 AA.
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Integumentary mucin B.1 (FIM-B.1) (Fragment).
 OS *Xenopus laevis* (African clawed frog).

AC Q96TL7; Q9UTW1; Q9UUL5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein rap1.
GN RAP1 OR SPBC1778.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
EX MEDLINE=21534324; PubMed=11676924;
RA Chikashige Y., Hiraoka Y.;
RT "Telomere binding of the Rap1 protein is required for meiosis in fission yeast.";
RT Curr. Biol. 11:1618-1623 (2001).
[2]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBUNIT.
EX MEDLINE=21534325; PubMed=11676925;
RA Kanoh J., Ishikawa F.;
RT "sprap1 and sprif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast.";
RT Curr. Biol. 11:1624-1630 (2001).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Horeby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Money P., Moule S., Murgall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volkart G., Aert R., Robben J., Grymonprez B., Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J.L., Moreno S., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Jerez J., Armstrong J., Forsburg S.L., Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe.";
RL Nature 415:871-880 (2002).
[4]
RP SEQUENCE OF 164-312 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T., Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190 (2000).
CC -!- FUNCTION: Involved in the regulation of telomere length, clustering and has a specific role in telomere position effect (TPE). Unlike yeast, exhibits now effect in transcription regulation.
CC -!- SUBUNIT: Interacts with taz1. Unlike yeast does not interact with rif1.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 BRCT domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch/announcements> or send an email to license@ebi.ac.uk).

EMBL; AB061738; BAB70735.1; -
EMBL; AY034032; AAK57740.1; -
DR EMBL; AL049489; CAC39280.1; -
DR EMBL; AB027870; BAA87174.1; -
DR EMBL; AB027963; BAA87267.1; -
DR GeneDB SPombe; SPBC1778.02; -
DR InterPro; IPR001357; BRCT.
DR PROSITE; PS01172; BRCT; 1.
KW DNA-binding; Telomere; Nuclear protein.
FT DOMAIN 14 105 BRCT.
SQ SEQUENCE 693 AA; 79526 MW; 77C5070697D1A9CF CRC64;

Query Match 37.6%; Score 47; DB 1; Length 693;
Best Local Similarity 47.4%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 YCQRLKRPQADCSAYTILS 20
||||:|
DB 187 YCQRIIVKPYSSQKDYIQS 205
|||:|

RESULT 8
PDXA_XYLFA
ID_PDXA_XYLFA STANDARD; PRT; 330 AA.
AC Q9PF39;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 4-hydroxythreonine-4-phosphate dehydrogenase (EC 1.1.1.262) (4-
DE (phosphohydroxy)-L-threonine dehydrogenase).
GN PDXA OR XF0839.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M., Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S., Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S., Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H., Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M., Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H., Faciniani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R., Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., Ho P.L., Hoesel J.D., Junqueira M.L., Kemp E.L., Kitajima J.P., Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.B., Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y., Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A., de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A., Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B., Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M., de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E., da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr., da Silva J.C.F., Silvestri M.I.Z., Siqueira W.J., de Souza A.A., de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H., Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

"The genome sequence of the plant pathogen *Xylella fastidiosa*.";
 RL Nature 406:151-159(2000).
 CC -!- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
 CC (phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-
 CC (phosphohydroxy)butyric acid which spontaneously decarboxylate to
 CC form 1-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl
 CC phosphate) (By similarity).
 CC -!- CATALYTIC ACTIVITY: 4-(phosphonoxy)-threonine + NAD(+) = 2-
 CC amino-3-oxo-4-phosphonoxybutyrate + NADH.
 CC -!- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
 CC pyridoxal phosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- SIMILARITY: Belongs to the pdxa family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003923; AAP83649.1; -;
 DR PIR; D82756; D82756.
 DR HAMAP; MF 00536; -; 1.
 DR InterPro; IPR005255; PdxA.
 DR Pfam; PF04166; PdxA; 1.
 DR TIGRfam; TIGR00557; pdxa; 1.
 KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
 SQ SEQUENCE 330 AA; 34553 MW; A981237FCE44CF71 CRC64;
 Query Match 36.8%; Score 46; DB 1; Length 330;
 Best Local Similarity 50.0%; Pred. No. 7.3;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BYCORVLRKPAQDC 14
 Db 19 ELCVRLVQPRQDC 32
 |||:::|||||
 RESULT 9
 PDXA_XYLFT STANDARD; PRT; 330 AA.
 ID PDXA_XYLFT
 AC Q87AJ1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 4-hydroxythreonine-4-phosphate dehydrogenase (EC 1.1.1.262) (4-
 OS (phosphohydroxy)-L-threonine dehydrogenase).
 GN PDXA OR PD1834.
 OS *Xylella fastidiosa* (strain Temecula / ATCC 700964).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; *Xylella*.
 OX NCBI_TaxID=183190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22421331; PubMed=12533478;
 RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
 RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
 RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
 RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
 RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
 RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
 RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
 RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
 RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
 RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,
 RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
 RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
 RA Kitajima J.P.;
 RT "Comparative analyses of the complete genome sequences of Pierce's
 RT disease and citrus variegated chlorosis strains of *Xylella*
 RT *fastidiosa*.";

J. Bacteriol. 185:1018-1026(2003).
 CC -!- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
 CC (phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-
 CC (phosphohydroxy)butyric acid which spontaneously decarboxylate to
 CC form 1-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl
 CC phosphate) (By similarity).
 CC -!- CATALYTIC ACTIVITY: 4-(phosphonoxy)-threonine + NAD(+) = 2-
 CC amino-3-oxo-4-phosphonoxybutyrate + NADH.
 CC -!- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
 CC pyridoxal phosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -!- SIMILARITY: Belongs to the pdxa family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE012560; AAO29666.1; -;
 DR HAMAP; MF 00536; -; 1.
 DR InterPro; IPR005255; PdxA.
 DR Pfam; PF04166; PdxA; 1.
 KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
 SQ SEQUENCE 330 AA; 34654 MW; 103FA63B9F0C1CF9 CRC64;
 Query Match 36.8%; Score 46; DB 1; Length 330;
 Best Local Similarity 50.0%; Pred. No. 7.3;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BYCORVLRKPAQDC 14
 Db 19 ELCVRLVQPRQDC 32
 |||:::|||||
 RESULT 10
 YLIE_ECOLI STANDARD; PRT; 782 AA.
 ID YLIE_ECOLI
 AC P75800;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein ylie.
 GN YLIE OR B0833.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RN Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi X.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 7.18-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RN DNA Res. 3:137-155(1996).


```

DR Pfam; PF00581; Rhodanese; 1.
DR PRINTS; PR00716; MIPHPHPTASE.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS0206; RHODANESE 3; 1.
KW Cell division; Mitosis; Hydrolyase; Multigene family.
FT DOMAIN 425 532 RHODANESE.
FT ACT SITE 481 481 BY SIMILARITY.
SQ SEQUENCE 574 AA; 64286 MW; 9367CE203B15FAAD CRC64;

Query Match 36.0%; Score 45; DB 1; Length 574;
Best Local Similarity 53.8%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Cc 3 CORVLRKPAQDCS 15
Cc 311 CQRLFRSPMPCS 323

RESULT 13
MPI2 MOUSE
ID MPI2_MOUSE STANDARD; PRT; 576 AA.
AC P30306; Q39LP3;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE M-phase inducer phosphatase 2 (EC 3.1.3.48) (Dual specificity
DE phosphatase Cdc25B).
GN CDC25B OR CDC25M2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=92217737; PubMed=1313771;
RX Kakizuka A., Sebastian B., Borgmeyer U., Hermans-Borgmeyer I.,
RA Bolado J., Hunter T., Hoekstra M.F., Evans R.M.;
RT "A mouse cdc25 homolog is differentially and developmentally
RT expressed."
RL Genes Dev. 6:578-590 (1992).
[2]
SEQUENCE OF 328-576 FROM N.A.
RP MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalon D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Fahy J.J., Helton E., Ketterman M., Madan A.C., Rodriques S., Sanchez A.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Functions as a dosage-dependent inducer in mitotic
CC control. It is a tyrosine protein phosphatase required for
CC progression of the cell cycle. It directly dephosphorylates CDC2
CC and activate its kinase activity.
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -!- ENZYME REGULATION: Stimulated by cyclins B.
CC -!- TISSUE SPECIFICITY: Expressed predominantly in spleen, lung,
CC heart, brain, intestine, and muscle.
CC -!- DEVELOPMENTAL STAGE: Detected at the one-cell stage followed by a

```

```

decrease in signal intensity at the two-cell stage. Detectable at
higher level in the four-cell stage and expressed through the
eight-cell, 16-cell and morula stages. Maximal expression at the
blastocyst stage.
-!- SIMILARITY: Belongs to the MPI phosphatase family.
-!- SIMILARITY: Contains 1 rhodanese domain.
---
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
---
EMBL; S93521; AAB22026.1; -.
EMBL; BC002287; AAB02287.1; -.
PIR; A42236; A42236.
HSSP; P30305; 1CWS.
MGD; MGI:99701; Cdc25b.
InterPro; IPR000751; MPI Phosphatase.
InterPro; IPR001763; Rhodanese-like.
Pfam; PF00581; Rhodanese; 1.
PRINTS; PR00716; MIPHPHPTASE.
SMART; SM00450; RHOD; 1.
PROSITE; PS0206; RHODANESE 3; 1.
Cell division; Mitosis; Hydrolyase; Multigene family.
DOMAIN 427 534 RHODANESE.
FT ACT SITE 483 483 BY SIMILARITY.
FT CONFLICT 532 532 F -> L (IN REF. 2).
SQ SEQUENCE 576 AA; 65490 MW; 5086B676581CD570 CRC64;

Query Match 36.0%; Score 45; DB 1; Length 576;
Best Local Similarity 53.8%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Cc 3 CORVLRKPAQDCS 15
Cc 313 CQRLFRSPMPCS 325

RESULT 14
MPI2 HUMAN
ID MPI2_HUMAN STANDARD; PRT; 580 AA.
AC P30305; Q43551; Q13971; Q9BRA6;
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE M-phase inducer phosphatase 2 (EC 3.1.3.48) (Dual specificity
DE phosphatase Cdc25B).
GN CDC25B OR CDC25H2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=92103683; PubMed=1836978;
RA Galaktionov K.I., Beach D.;
RT "Specific activation of cdc25 tyrosine phosphatases by B-type
RT cyclins: evidence for multiple roles of mitotic cyclins."
RL Cell 67:1181-1194 (1991).
[2]
SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=92118716; PubMed=1662986;
RA Nagata A., Igarashi M., Jinno S., Suto K., Okayama H.;
RT "An additional homolog of the fission yeast cdc25+ gene occurs in
RT humans and is highly expressed in some cancer cells."
RL New Biol. 3:959-968 (1991).
[3]
SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RX MEDLINE=97332629; PubMed=9188863;
RA Baldwin V., Cans C., Superti-Purga G., Docomun B.;

```

RT "Alternative splicing of the human CDC25B tyrosine phosphatase.
 RT Possible implications for growth control?";
 RL Oncogene 14:2485-2495 (1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beare L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahm D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Hartley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Lehaealainho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McIlroy K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20";
 RL Nature 414:865-871 (2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Hellon E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [6]
 RP SEQUENCE OF 56-352 FROM N.A. (ISOFORMS 2 AND 3).
 RX McCormack A.K., DeSouza C.C.P.C., Tonks I.D., Clark J.M.,
 RA Forrest A.R.R., Hayward N.K., Ellem K.A.O., Gabrielli B.G.;
 RT "Alternative splicing of cdc25B";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (1.91 ÅNGSTRÖMS) OF 370-580.
 RX MEDLINE=20013068; PubMed=10543950;
 RA Reynolds R.A., Yem A.W., Wolfe C.L., Deibel M.R. Jr., Chidester C.G.,
 RA Watenpaugh K.D.;
 RT "Crystal structure of the catalytic subunit of Cdc25B required for
 G2/M phase transition of the cell cycle";
 RL J. Mol. Biol. 293:559-568 (1999).
 CC -!- FUNCTION: Functions as a dosage-dependent inducer in mitotic
 control. It is a tyrosine protein phosphatase required for

CC progression of the cell cycle. It directly dephosphorylates CDC2
 and activate its kinase activity. The three isoforms seem to have
 a different level of activity.
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
 tyrosine + phosphate.
 CC -!- ENZYME REGULATION: Stimulated by cyclins B.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=3; Synonyms=CDC25B3;
 CC IsoId=P30305-1; Sequence=Displayed;
 CC Name=1; Synonyms=CDC25B1;
 CC IsoId=P30305-2; Sequence=VSP_000861;
 CC Name=2; Synonyms=CDC25B2;
 CC IsoId=P30305-3; Sequence=VSP_000862;
 CC -!- SIMILARITY: Belongs to the MPI phosphatase family.
 CC -!- SIMILARITY: Contains 1 rhodanese domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81934; AAA58416.1; -
 CC EMBL; S78187; AAB21139.1; -
 CC EMBL; Z68092; CAA92108.1; -
 CC EMBL; X96436; CAA65303.1; -
 CC EMBL; BC051711; AAH51711.1; -
 CC EMBL; AL109804; CAC17549.1; -
 CC EMBL; AL109804; CAC17548.1; -
 CC EMBL; AL109804; CAC32458.1; -
 CC EMBL; SC006395; AAH06395.1; -
 CC EMBL; BC009953; AAH09953.1; -
 CC EMBL; AF036233; AAB94622.1; -
 CC EMBL; AF036233; AAB94624.1; -
 CC PIR; B41648; B41648.
 CC PDB; 1Q80; 29-APR-00.
 CC PDB; 1CWR; 28-AUG-00.
 CC PDB; 1CWS; 30-AUG-00.
 CC PDB; 1CWT; 30-AUG-00.
 CC Genew; HGNC:1726; CDC25B.
 CC GK; P30305; -
 CC MIM; 116949; -
 CC GO; GO:0005622; C:intracellular; NAS.
 CC GO; GO:0004725; P:protein tyrosine phosphatase activity; TAS.
 CC GO; GO:0007067; P:mitosis; TAS.
 CC GO; GO:0000278; P:mitotic cell cycle; TAS.
 CC GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
 CC GO; GO:0000074; P:regulation of cell cycle; TAS.
 CC InterPro; IPR000751; MPI Phosphatase.
 CC InterPro; IPR001763; Rhodanese-like.
 CC Pfam; PF00581; Rhodanese; 1.
 CC SMART; SM00450; RHOD; 1.
 CC PROSITE; PS50206; RHODANESE 3; 1.
 KW Cell division; Mitosis; Hydrolase; Alternative splicing;
 KW Multigene family; 3D-structure.
 FT DOMAIN 431 538
 FT ACT_SITE 487 487
 FT VARSPLIC 68 81
 FT VARSPLIC 154 194
 FT CONFLICT 575 575
 FT SEQUENCE 580 AA; 64987 MW; EDE24H0E84AC1BE3 CRC64;
 SQ
 Query Match 36.0%; Score 45; DB 1; Length 580;
 Best Local Similarity 53.8%; Pred. No. 19;
 Matches 7; Conservative 2; Mismatches 4; Indels 0;
 Gaps 0;
 QY 3 CORVLRKPAQDCS 15

Db 315 CQRLFRSPMPCS 327

||||: ||: ||

RESULT 15

VP35 SCHPO

ID VP35 SCHPO STANDARD; PRT; 785 AA.

AC O74552; P78930;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Vacuolar protein sorting-associated protein vps35.

GN VPS35 OR SPCC777.13.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

SEQUENCE FROM N.A.

RP STRAIN=972;

RC MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volkheart G., Aert R., Robben J., Grymonprez B.,

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.A., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,

RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.";

RN Nature 415:871-880(2002).

RN [2]

SEQUENCE OF 279-785 FROM N.A.

RP STRAIN=PR745;

RC MEDLINE=98162722; PubMed=9501991;

RX Yoshioka S., Kato K., Nakai K., Okayama H., Nojima H.;

RA "Identification of open reading frames in Schizosaccharomyces pombe

RT cDNAs.";

RL DNA Res. 4:363-369(1997).

CC -!- FUNCTION: May play a role in vesicular protein sorting, analogous

CC to the yeast retromer proteins (By similarity).

CC -!- SUBUNIT: Probably a component of a membrane-associated

CC multiprotein complex (By similarity).

CC -!- SUBCELLULAR LOCATION: Associated with an intracellular membrane

CC (By similarity).

CC -!- SIMILARITY: Belongs to the Vps35 family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch).

CC -----

CC RREMBL; AL031532; CAA20717.1; -

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:51:12 ; Search time 3.46927 Seconds
(without alignments)
2091.769 Million cell updates/sec

Title: US-09-980-403-2_COPY_804_826
Perfect score: 125
Sequence: 1 EYQCVLRKPAQDCSAVTLSDFT 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	125	100.0	823	4 Q8WY18	Q8WY18 homo sapien
2	112	89.6	823	11 Q8CE84	Q8CE84 mus musculus
3	112	89.6	1188	11 Q7TQC3	Q7TQC3 mus musculus
4	57	45.6	300	16 Q89GR9	Q89GR9 bradyrhizob
5	52	41.6	198	5 Q22857	Q22857 caenorhabdi
6	51.5	41.2	897	10 Q7XSB7	Q7XSB7 oryza sativ
7	50	40.0	86	5 Q8260	Q8260 drosophila
8	49	39.2	148	16 Q9KV41	Q9KV41 fowipox chol
9	49	39.2	1949	12 Q9J5C1	Q9J5C1 fowipox vir
10	48	38.4	1184	4 Q75J39	Q75J39 homo sapien
11	48	38.4	1184	4 Q8IVY5	Q8IVY5 homo sapien
12	48	38.4	1250	11 Q8BV01	Q8BV01 mus musculus
13	47	37.6	335	5 Q9UD3	Q9UD3 caenorhabdi
14	47	37.6	387	5 Q18174	Q18174 caenorhabdi
15	47	37.6	898	5 Q86JG0	Q86JG0 dictyosteli
16	46.5	37.2	1013	10 Q7XRH7	Q7XRH7 oryza sativ

17	46.5	37.2	1436	11	Q35564	Q35564 mus musculus
18	46.5	37.2	1436	11	P70125	P70125 mus musculus
19	46	36.8	223	16	O84565	O84565 chlamydia t
20	46	36.8	235	16	Q9PUH9	Q9PUH9 chlamydia t
21	46	36.8	254	10	Q9AVH8	Q9AVH8 pismu sativ
22	46	36.8	329	13	Q8JIZ9	Q8JIZ9 brachydanio
23	46	36.8	572	5	O17123	O17123 caenorhabdi
24	46	36.8	622	5	Q9VWP5	Q9VWP5 drosophila
25	46	36.8	782	16	Q8X6V4	Q8X6V4 escherichia
26	46	36.8	782	16	Q8FJK7	Q8FJK7 escherichia
27	46	36.8	1122	11	Q7TT33	Q7TT33 mus musculus
28	46	36.8	1275	3	Q9Y7U5	Q9Y7U5 schizosacch
29	46	36.8	2203	5	Q963L8	Q963L8 schistosoma
30	46	36.8	2731	5	Q9VJTS	Q9VJTS drosophila
31	46	36.8	3367	5	Q9XZC9	Q9XZC9 drosophila
32	46	36.8	3375	5	Q8IP51	Q8IP51 drosophila
33	45.5	36.4	418	2	Q8GMF6	Q8GMF6 streptomyce
34	45	36.0	77	4	Q9HIH5	Q9HIH5 homo sapien
35	45	36.0	99	3	Q8X015	Q8X015 neurospora
36	45	36.0	165	16	Q8ENAL	Q8ENAL oceanobacil
37	45	36.0	217	4	Q9HGF0	Q9HGF0 homo sapien
38	45	36.0	270	11	Q8BV50	Q8BV50 mus musculus
39	45	36.0	305	4	O43550	O43550 homo sapien
40	45	36.0	330	16	Q89FV0	Q89FV0 bradyrhizob
41	45	36.0	396	2	Q93L34	Q93L34 rhizobium l
42	45	36.0	431	11	Q8BYA5	Q8BYA5 mus musculus
43	45	36.0	468	6	Q9SRE6	Q9SRE6 macaca fasc
44	45	36.0	477	4	Q8TEB8	Q8TEB8 homo sapien
45	45	36.0	531	11	Q8BWI8	Q8BWI8 mus musculus

ALIGNMENTS

RESULT 1

ID	Q8WY18	PRELIMINARY;	PRT;	823 AA.
AC	Q8WY18;			
DT	01-MAR-2002	(TRENBLrel. 20, Created)		
DT	01-MAR-2002	(TRENBLrel. 20, Last sequence update)		
DT	01-JUN-2003	(TRENBLrel. 24, Last annotation update)		
DE	MSTP018.			
GN	MST018.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Aorta;			
RA	Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,			
RA	Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,			
RA	Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,			
RA	Gao R.L., Qiang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF111799; AAL39001.1; -			
DR	GO: GO:0008305; C:integrin complex; IEA.			
DR	GO: GO:0004895; F:cell adhesion receptor activity; IEA.			
DR	GO: GO:0007160; P:cell-matrix adhesion; IEA.			
DR	InterPro: IPR000413; Integrin_alpha.			
DR	Pfam: PF01839; FG-GAP; 3.			
DR	PRINTS: PR01185; INTEGRINA.			
DR	SMART: SM00191; Int_alpha; 4.			
SQ	SEQUENCE 823 AA; 92672 MW; DE4E78079DCD4925 CRC64;			

Query Match 100.0%; Score 125; DB 4; Length 823;

Best Local Similarity 100.0%; Pred No. 5.3e-12; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQCVLRKPAQDCSAVTLSDFT 23

|||||

Db 439 EYQCVLRKPAQDCSAVTLSDFT 461

```
RESULT 2
Q8CE84 PRELIMINARY; PRT; 823 AA.
ID Q8CE84;
AC Q8CE84;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE MSTP018 homolog.
GN 4732459H24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).
DR EMBL; AK028821; BAC26137.1; -.
DR GO; MGI:2442114; 4732459H24RIK.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; F:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 3.
DR PRINTS; PR01185; INTEGRIN.
DR SMART; SM00191; Int_alpha; 4.
DR SEQUENCE 823 AA; 92284 MW; A330236324A0E089 CRC64;
SQ
Query Match 89.6%; Score 112; DB 11; Length 823;
Best Local Similarity 87.0%; Pred. No. 7.6e-10;
Matches 20; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYCORVLRKPAQDCSAYTSLSFT 23
Db 439 EYCORVLRKPAQDCSAYTSLSFT 461
|||||:|||||:|||||

RESULT 3
Q7TQC3 PRELIMINARY; PRT; 1188 AA.
ID Q7TQC3
AC Q7TQC3;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE A11 integrin.
GN ITG11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Tiger C.-F., Popova S.N., Velling T., Ekman S., Forsberg E.,
RA Gullberg D.;
RT "A11 integrin is important for mesenchymal cell function:
RT elimination of a1b1 leads to dwarfism.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Johansson M., Popova S.N.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV124460; AAM62130.1; -.
KW Integrin.
SQ SEQUENCE 1188 AA; 133067 MW; 69B2BF20D75E4271 CRC64;

Query Match 89.6%; Score 112; DB 11; Length 1188;
Best Local Similarity 87.0%; Pred. No. 1.1e-09;
Matches 20; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
Matches 20; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EYCORVLRKPAQDCSAYTSLSFT 23
Db 804 EYCORVLRKPAQDCSAYTSLSFT 826
|||||:|||||:|||||

RESULT 4
Q89GR9 PRELIMINARY; PRT; 300 AA.
ID Q89GR9
AC Q89GR9;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE B116276 protein.
GN B116276.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
DR EMBL; AP005958; BAC51541.1; -.
DR InterPro; IPR004360; Gly_bleo_diox.
DR Pfam; PF00903; Glyoxalase; 2.
KW Complete proteome.
SQ SEQUENCE 300 AA; 33603 MW; 145AB3D7ADD4B11E CRC64;

Query Match 45.6%; Score 57; DB 16; Length 300;
Best Local Similarity 47.6%; Pred. No. 0.38;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 QYRVLRKPAQDCSAYTSLSFT 23
Db 87 CRVSRKPSLDCMAAGVTFT 107
|||||:|||||:|||||

RESULT 5
Q22857 PRELIMINARY; PRT; 198 AA.
ID Q22857
AC Q22857;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE T28H10.2 protein.
GN T28H10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z75551; CAA99934.1; -.
DR PIR; T25436; T25436.
DR WormPep; T28H10.2; CE06530.
```

```

DR InterPro; IPR003582; SHKT.
DR Pfam; PF01549; SHKT; 1.
DR SMART; SMC0254; SHKT; 1.
SQ SEQUENCE 198 AA; 22741 MW; BCD5CCAB905EBE5 CRC64;

Query Match 41.6%; Score 52; DB 5; Length 198;
Best Local Similarity 43.5%; Pred. No. 1.7;
Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 EYQCVLRKPAQDCSAYTLSPDT 23
   |||::|||:|:|:|:|
Db 49 EYIKKCKKSCGNCPRYELKFPDT 71

RESULT 6
Q7XS87
ID Q7XS87 PRELIMINARY; PRT; 897 AA.
AC Q7XS87
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE QJ990528 30.6 protein.
GN QJ990528_30.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
RA Jia G., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
RA Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
RA Chen L., Fan D.B., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
RA Lu T.F., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
RA Zhang R.Q., Guan J.P., Hong G.F.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL662938; CAF02048.1; -.
DR EMBL; AL662938; CAF02048.1; -.
SQ SEQUENCE 897 AA; 100727 MW; B43ADF8E556CBB22 CRC64;

Query Match 41.2%; Score 51.5; DB 10; Length 897;
Best Local Similarity 57.1%; Pred. No. 9.3;
Matches 12; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

QY 1 EYQCVLRKPAQDCSAYTLSP 21
   |||:|:|:|:|:|:|
Db 601 EYCPRTLRL--QDPETYSGF 618

RESULT 7
Q08260
ID Q08260 PRELIMINARY; PRT; 86 AA.
AC Q08260;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE JINGWEI protein (Fragment).
GN JGW.
OS Drosophila teissieri (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7243;
RN [1]
RP SEQUENCE FROM N.A.
RA Long M., Langley C.H.;
RA MEDLINE=93219840; PubMed=7682012;
RT "Natural selection and the origin of jingwei, a chimeric processed
functional gene in Drosophila.";
RL Science 260:91-95(1993).

```

```

DR EMBL; S57937; AAD13889.1; -.
DR FlyBase; FBgn0013013; Dtel\jgw.
FT NON_TER 86
SQ SEQUENCE 86 AA; 9424 MW; C73C9269418B695C CRC64;

Query Match 40.0%; Score 50; DB 5; Length 86;
Best Local Similarity 50.0%; Pred. No. 1.6;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 QCVLRKPAQDCSAYTL 20
   ||:|:|:|:|:|
Db 65 COLQFSRPASKCAFTLS 82

RESULT 8
Q9KV41
ID Q9KV41 PRELIMINARY; PRT; 148 AA.
AC Q9KV41;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein VC0317.
GN VC0317.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483 (2000).
DR EMBL; AE004120; AAF93490.1; -.
DR PIR; C82337; C82337.
DR TIGR; VC0317; -.
DR GO; GO:0008080; P:N-acetyltransferase activity; IEA.
DR InterPro; IPR000182; GCNSacetyl_trans.
DR Pfam; PF00583; Acetyltransf_1.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 148 AA; 17392 MW; 472417CDDCAFE572 CRC64;

Query Match 39.2%; Score 49; DB 16; Length 148;
Best Local Similarity 50.0%; Pred. No. 3.9;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 EYQCVLRKPAQDCSAYT 18
   :|:|:|:|:|:|
Db 86 DYQCKQWLNPATFCPAYT 103

RESULT 9
Q9J5C1
ID Q9J5C1 PRELIMINARY; PRT; 1949 AA.
AC Q9J5C1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ORF FPV099 variola B22R gene family protein.
GN FPV099.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]

```



```

DE Cartilage intermediate layer protein homolog.
GN CILP OR C33036G17RIX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK081544; BAC38252.1; -.
DR MGD; MGI:244507; Cilp.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG c2.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00090; csp_1; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS0835; IG LIKE; 1.
DR PROSITE; PS50092; TSP1; 1.
SQ SEQUENCE 1250 AA; 139709 MW; A059B1A3284DB475 CRC64;

Query Match 38.4%; Score 48; DB 11; Length 1250;
Best Local Similarity 40.0%; Pred. No. 49;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 2 YCQVLRKPAQDCSAYTLSF 21
Db 118 WCLNREQRGQNCNRYTRF 137

RESULT 13
Q9U4D3 PRELIMINARY; PRT; 335 AA.
ID Q9U4D3
AC Q9U4D3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative tyrosine kinase receptor W04G5.6A (Fragment).
GN W04G5.6A.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=20036932; PubMed=10568743;
RA Popovici C., Roulin R., Coulter P., Pontarotti P., Birnbaum D.;
RT "The family of Caenorhabditis elegans tyrosine kinase receptors:
RT similarities and differences with mammalian receptors.";
RL Genome Res. 9:1026-1039 (1999).
DR EMBL; AF188749; AAF00546.1; -.
DR PIR; T37262; T37262.
DR HSSP; P11362; IFGK.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 387 AA; 44311 MW; F851FF45234171A4 CRC64;

Query Match 37.6%; Score 47; DB 5; Length 387;
Best Local Similarity 38.1%; Pred. No. 22;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 2 YCQVLRKPAQDCSAYTLSF 22
Db 91 YRAKLRKPTQEVSSHIVFD 111

RESULT 14
O18174 PRELIMINARY; PRT; 387 AA.
ID O18174
AC O18174;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE W04G5.6 protein.
GN W04G5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z93391; CAB07686.3; -.
DR PIR; T26176; T26176.
DR HSSP; P11362; IFGK.
DR WormPep; W04G5.6; CE28099.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 387 AA; 44311 MW; F851FF45234171A4 CRC64;

Query Match 37.6%; Score 47; DB 5; Length 387;
Best Local Similarity 38.1%; Pred. No. 22;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 2 YCQVLRKPAQDCSAYTLSF 22
Db 91 YRAKLRKPTQEVSSHIVFD 111

```

RESULT 15

Q86JG0 PRELIMINARY; PRT; 898 AA.

ID Q86JG0
AC Q86JG0;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to Dictyostelium discoideum (Slime mold). LagC2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116982; AAO51616.1; --
DR InterPro; IPR007110; I9-like.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF01833; TIG; 3.
SQ SEQUENCE 898 AA; 97895 MW; 54EEAB78E636E341 CRC64;

Query Match 37.6%; Score 47; DB 5; Length 898;
Best Local Similarity 47.1%; Pred. No. 52;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 CORVLRKPAQCGSAYTL 19
DB 89 CAREEADPRDCTSYTL 105

Search completed: September 21, 2004, 13:03:48
Job time : 6.46927 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:43:42 : Search time 4.38478 Seconds
(without alignments)
1482.081 Million cell updates/sec

Title: US-09-980-403-2_COPY_804_826
Perfect score: 125
Sequence: 1 EYQVRLKXPAQDCSAVTLSPDT 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	125	100.0	1034	3	AAB25590 Protein e
2	125	100.0	1034	6	ADA27062 Human nov
3	125	100.0	1034	8	ADE86592 Novel hum
4	125	100.0	1120	6	ABR58365 Human NOV
5	125	100.0	1188	4	AAB30929 Amino aci
6	125	100.0	1188	4	AAU14467 Human nov
7	125	100.0	1188	4	AAU14231 Human nov
8	125	100.0	1188	4	AAB50085 Human A25
9	125	100.0	1188	5	AAU10551 Human A25
10	125	100.0	1188	7	ADE09956 Novel pro
11	125	100.0	1189	3	AAB25582 ITGAl1 pr
12	125	100.0	1189	4	ABG12949 Novel hum
13	125	100.0	1189	6	ABR58364 Human NOV
14	125	100.0	1189	6	ADA27054 Human nov
15	125	100.0	1189	7	ADE63570 Human Pro
16	125	100.0	1189	8	ADE86584 Novel hum
17	115	92.0	545	5	ABB72288 Murine pr
18	115	92.0	688	5	ABB72300 Rat prote
19	115	92.0	696	5	ABB72289 Rat prote
20	112	89.6	1188	4	AAB50087 Murine A2
21	112	89.6	1188	5	AAU10552 Murine A2
22	48	38.4	1184	2	AAW74445 Human nuc
23	48	38.4	1184	3	AAV66657 Membrane-
24	48	38.4	1184	4	AAU12377 Human PRO
25	48	38.4	1184	4	AAB65180 Human PRO

26	48	38.4	1184	6	ABU57995	Human PRO
27	48	38.4	1184	6	ABU59073	Novel hum
28	48	38.4	1184	6	ABU82585	Human sec
29	48	38.4	1184	6	ABO17821	Novel hum
30	48	38.4	1184	6	ABU60504	Human sec
31	48	38.4	1184	6	ABU13886	Human PRO
32	48	38.4	1184	6	ABU81075	Human PRO
33	48	38.4	1184	6	ABU72471	Novel hum
34	48	38.4	1184	6	ABU66775	Human PRO
35	48	38.4	1184	6	ABU59856	Novel sec
36	48	38.4	1184	6	ABU59220	Human sec
37	48	38.4	1184	6	ABO25917	Human PRO
38	48	38.4	1184	6	ABO25046	Human sec
39	48	38.4	1184	6	ABU58926	Human sec
40	48	38.4	1184	6	ABU92304	Novel hum
41	48	38.4	1184	6	ABU59369	Novel hum
42	48	38.4	1184	6	ABU67051	Human sec
43	48	38.4	1184	6	ABU92135	Novel hum
44	48	38.4	1184	6	ABU10841	Human PRO
45	48	38.4	1184	6	ABU81593	Novel hum

ALIGNMENTS

RESULT 1
AAB25590
ID AAB25590 standard; protein; 1034 AA.
XX AC AAB25590;
XX XX

DT 21-NOV-2000 (first entry)
XX XX

DE Protein encoded by human secreted protein gene #7 clone HOHBY69.
XX XX

KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
KW antirheumatic, dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnery; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
XX XX

OS Homo sapiens.
XX XX

FN WO200029435-A1.
XX XX

PD 25-MAY-2000.
XX XX

PF 27-OCT-1999; 99WO-US025031.
XX XX

PR 28-OCT-1998; 98US-0105971P.
XX XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX XX

PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
PI Greene JW;
XX XX

WPI; 2000-387742/33.
XX XX

PT Isolated nucleic acid molecules encoding human secreted proteins are used
PT for the prevention, amelioration and treatment of autoimmune,
PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
XX wounds, and infectious diseases.
XX XX

PS Claim 1; Page 678-682; 803pp; English.
XX XX

CC The present invention relates to 12 secreted human proteins and the
CC nucleotide sequences encoding them. The polynucleotide sequences given in
CC AAA80606-A80623 encode the 12 secreted protein sequences given in
CC AAB25576-B25593. The human secreted proteins have various activities
CC dependent on the tissues in which they are expressed. Examples of the

CC activities of the proteins include: immunosuppressant; anti-inflammatory;
 CC antiarthritic; antirheumatic, dermatological; antiproliferative;
 CC antiarteriosclerotic; anticancer; vulnary; antiviral; antibacterial;
 CC and antifungal activity. The proteins, polypeptides, agonists and
 CC antagonists may be used to treat prevent and/or diagnose various disease,
 CC disorders and conditions examples of which include: immune disorders e.g.
 CC Addison's disease, rheumatoid arthritis, dermatitis, and multiple
 CC sclerosis; inflammatory disorders e.g. inflammatory bowel disease,
 CC Crohn's disease and nephritis; hyperproliferative disorders such as
 CC paraproteinemia and purpura; cardiovascular disorders e.g. coronary
 CC arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The
 CC proteins and polynucleotide sequences may also be used in wound healing
 CC and the treatment of infectious diseases. The human secreted protein gene
 CC #7 and protein sequences are represented in sequences AAA80612 and
 CC AAA25582. Secreted protein gene #7 is located at position chromosome 15
 CC q22.3-23. Sequences AAA80652-AA80661 represent genes which are related to
 CC the secreted protein gene#7
 XX
 SQ Sequence 1034 AA;

Query Match 100.0%; Score 125; DB 3; Length 1034;
 Best Local Similarity 100.0%; Pred. No. 6.9e-11;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQVRLKPKAQCDSAYTILSPT 23
 DB 804 EYQVRLKPKAQCDSAYTILSPT 826

RESULT 2

ADA27062
 ID ADA27062 standard; protein; 1034 AA.

AC ADA27062;

DT 20-NOV-2003 (first entry)

DE Human novel secreted protein from cDNA HOBY69 #2.

XX cytostatic; antiinflammatory; immunomodulator; neuroprotective;
 KW hemostatic; gene therapy; cancer; inflammation; immune disorder;
 KW neurological disorder; blood clotting disorder; food additive;
 KW preservative; human; secreted protein.

OS Homo sapiens.

XX US2003055231-A1.

PN 20-MAR-2003.

PD 29-OCT-2001; 2001US-00984130.

PF 28-OCT-1998; 98US-0105971P.

PR 27-OCT-1999; 99WO-US025031.

PR 19-APR-2000; 2000US-0198407P.

PR 30-OCT-2000; 2000US-0243792P.

PR 18-APR-2001; 2001US-00836353.

XX (NIJ/) NI J.

PA (YOON/) YOUNG P E.

PA (KENN/) KENNY J J.

PA (OLSE/) OLSEN H S.

PA (MOOR/) MOORE P A.

PA (WEI/) WEI Y.

PA (GREE/) GREENE J M.

PA (RUBE/) RUBEN S M.

PA (LIUD/) LIU D.

PA (CROC/) CROCKER P R.

XX Ni J, Young PE, Kenny JJ, Olsen HS, Moore PA, Wei Y, Greene JM;

PI Ruben SM, Liu D, Crocker PR;

XX WPI; 2003-567103/53.

DR

DR N-PSDB; ADA27044.

XX New human secreted nucleic acid molecules and polypeptides, useful for
 PT preventing, treating, or ameliorating a medical condition, such as
 PT cancer, inflammation, immune disorders, neurological and blood clotting
 PT disorders.

XX Claim 11; Page 302-305; 454pp; English.

XX The invention relates to an isolated nucleic molecule that is at least
 CC 95% identical to 18 human cDNA sequences representing 12 novel genes
 CC encoding secreted proteins or a polynucleotide fragment of the cDNA
 CC sequence contained in American Type Culture Collection (ATCC) deposit No.
 CC defined in the specification, its species homologue, a variant or allelic
 CC variant of the polynucleotide having a polynucleotide capable of
 CC hybridising under conditions the polynucleotide, where the polynucleotide
 CC does not hybridise under stringent conditions to a nucleic acid molecule
 CC having a nucleotide sequence of only A or T residues. Also included are
 CC recombinant vectors, host cells (for producing the polypeptide), the
 CC secreted polypeptide (comprising a sequence that is at least 95%
 CC identical to a polypeptide fragment, domain, epitope, full-length
 CC protein, variant, allelic variant or species homologue), antibodies that
 CC specifically bind to the polypeptides, diagnosing, treating, preventing
 CC or ameliorating a medical condition by administering the polynucleotide
 CC or the polypeptide, the gene corresponding to the cDNA sequence and
 CC identifying an activity in a biological assay (by expressing the cDNA
 CC sequence in a cell, isolating the supernatant, and detecting an activity
 CC in a biological assay and identifying the protein in the supernatant
 CC having the activity). The polypeptides, nucleic acids and antibodies are
 CC useful for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition, for preventing, treating, or ameliorating a
 CC medical condition, such as cancer, inflammation and other immune
 CC disorders, neurological and blood clotting disorders (many examples are
 CC given in the specification). The nucleic acids are also useful for
 CC chromosome identification, radiation hybrid mapping or long-range
 CC restriction mapping. The polypeptides and antibodies are useful for
 CC providing immunological probes for differential identification of the
 CC tissues immunohistochemistry assays. The polypeptide, polynucleotide,
 CC agonist or antagonist may also be used as a food additive or preservative
 CC to increase or decrease storage capabilities, fat content or other
 CC nutritional components. The present is a secreted protein of the
 CC invention.

XX Sequence 1034 AA;

Query Match 100.0%; Score 125; DB 6; Length 1034;
 Best Local Similarity 100.0%; Pred. No. 6.9e-11;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQVRLKPKAQCDSAYTILSPT 23
 DB 804 EYQVRLKPKAQCDSAYTILSPT 826

RESULT 3

ADE86592
 ID ADE86592 standard; protein; 1034 AA.

XX ADE86592;

DT 29-JAN-2004 (first entry)

DE Novel human secreted protein #15.

XX human; secreted protein; cancer; liver disorder; hepatitis;
 KW neural disorder; Alzheimer's disease.

XX Homo sapiens.

XX US2003129685-A1.

XX 10-JUL-2003.

XX

```

PF 18-APR-2001; 2001US-00836353.
XX
PR 28-OCT-1998; 98US-0105971P.
PR 27-OCT-1999; 99WO-US025031.
PR 19-APR-2000; 2000US-0198407P.
XX
XX (NIJUN/) NI J.
PA (YOUNG/) YOUNG P E.
PA (KENN/) KENNY J J.
PA (OLSEN/) OLSEN H S.
PA (MOORE/) MOORE P A.
PA (WEIY/) WEI Y.
PA (GREEE/) GREENE J M.
PA (RUBE/) RUBEN S M.
XX
PI Ni J, Young PE, Kenny JJ, Olsen HS, Moore PA, Wei Y, Greene JM;
PI Ruben SM;
XX
XX WPI; 2004-020335/02.
DR N-PSDB; ADE86574.
XX
XX New nucleic acid molecule, useful for preparing a medicament for
PT preventing, treating or ameliorating a medical condition e.g. cancer,
PT liver disorders or neural disorders.
XX
XX Claim 11; SEQ ID NO 43; 380pp; English.
XX
XX The invention relates to an isolated nucleic acid sequence, or its
CC allelic variant, a fragment of the cDNA sequence, or its fragment,
CC domain, epitope or species homologue. The nucleic acid is useful for
CC preparing a medicament for preventing, treating or ameliorating a medical
CC condition e.g., cancer, liver disorders such as hepatitis or neural
CC disorders such as Alzheimer's disease. The present sequence represents
CC the amino acid sequence of a novel human secreted protein.
XX
XX Sequence 1034 AA;
XX
XX Query Match 100.0%; Score 125; DB 8; Length 1034;
XX Best Local Similarity 100.0%; Pred. No. 6.9e-11;
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 EYCVLRKPAQDCSAYTLSPDT 23
XX |||||
Db 804 EYCVLRKPAQDCSAYTLSPDT 826
XX
XX RESULT 4
XX ABR58365
XX AC ABR58365;
XX
XX 07-JUL-2003 (first entry)
XX
XX Human NOV2b.
XX
XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;
XX immunomodulator; cytosratic; nootropic; neuroprotective; dyslipidaemia;
XX antiparkinsonian; antilipemic; gene therapy; metabolic disorder;
XX diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;
XX neurodegenerative disorder; Alzheimer's disease; immune disorder;
XX haematopoietic disorder.
XX
XX Homo sapiens.
XX
XX WO2003029423-A2.
XX
XX 10-APR-2003.
XX
XX 02-OCT-2002; 2002WO-US031358.
XX
XX 02-OCT-2001; 2001US-0326483P.
PR 05-OCT-2001; 2001US-0327342P.

```

```

PR 09-OCT-2001; 2001US-0327917P.
PR 09-OCT-2001; 2001US-0328029P.
PR 09-OCT-2001; 2001US-0328044P.
PR 09-OCT-2001; 2001US-0328056P.
PR 12-OCT-2001; 2001US-0328849P.
PR 15-OCT-2001; 2001US-0329414P.
PR 17-OCT-2001; 2001US-0330142P.
PR 22-OCT-2001; 2001US-0341088P.
PR 24-OCT-2001; 2001US-0339266P.
PR 24-OCT-2001; 2001US-0343629P.
PR 29-OCT-2001; 2001US-0349575P.
PR 01-NOV-2001; 2001US-0346357P.
PR 12-APR-2002; 2002US-0371972P.
PR 12-APR-2002; 2002US-0371980P.
PR 17-APR-2002; 2002US-0373261P.
PR 19-APR-2002; 2002US-0373805P.
PR 23-APR-2002; 2002US-0374738P.
PR 16-MAY-2002; 2002US-0381101P.
PR 17-MAY-2002; 2002US-0381635P.
PR 29-MAY-2002; 2002US-0383830P.
PR 01-OCT-2002; 2002US-00262839.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;
PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;
PI Rothenberg ME, Shinkets RA, Smithson G, Spytek KA, Taupier RJ;
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;
XX
XX WPI; 2003-381625/36.
XX N-PSDB; ACC72077.
XX
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or
PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or
PT dyslipidemia, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
XX Claim 1; Page 107; 487pp; English.
XX
XX The present invention relates to novel human NOV proteins and their
CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV
CC proteins are useful in manufacturing a medicament for treating a syndrome
CC associated with a human disease. The NOV proteins and coding sequences
CC may be used to diagnose, treat or prevent metabolic disorders such as
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative
CC disorders such as Alzheimer's disease or Parkinson's disease, immune
CC disorders, haematopoietic disorders and various dyslipidaemias
XX
XX Sequence 1120 AA;
XX
XX Query Match 100.0%; Score 125; DB 6; Length 1120;
XX Best Local Similarity 100.0%; Pred. No. 7.5e-11;
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 EYCVLRKPAQDCSAYTLSPDT 23
XX |||||
Db 735 EYCVLRKPAQDCSAYTLSPDT 757
XX
XX RESULT 5
XX AAB30929
XX ID AAB30929 standard; protein; 1188 AA.
XX
XX AAB30929;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a human alpha11 integrin chain.
XX
XX Human; integrin; alpha11 subunit; fibroblast; muscle cell; chondrocyte;
XX osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;
XX wound healing; trauma; rheumatoid arthritis; osteoarthritis;
XX

```

KW osteoporosis; cartilage damage; bone damage; cartilage.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..22 /note= "signal peptide"
 FT Region 951..972 /note= "leucine zipper"
 FT Domain 1142..1164 /note= "transmembrane domain"
 FT
 XX
 XX WO200075187-A1.
 XX
 XX PD 14-DEC-2000.
 XX
 XX PF 31-MAY-2000; 2000WO-SE001135.
 XX
 XX PR 03-JUN-1999; 95SE-00002056.
 XX
 XX PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 XX PI Gullberg D;
 XX
 XX DR WPI; 2001-071061/08.
 XX DR N-PSDB; AAC86971.
 XX
 XX Integrin subunit alpha 11 or integrin heterodimer comprising subunit
 PT alpha 11 in association with subunit beta, useful for treating muscle
 PT dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis.
 PT
 XX
 XX PS Disclosure; Fig 2a-c; 79pp; English.
 XX
 CC The present sequence represents a human integrin subunit, designated
 CC alpha11. The alpha11 polynucleotide and polypeptide are useful as markers
 CC of cell target molecules, such as fibroblasts, muscle cells,
 CC chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.
 CC They are also used for determining the differential-stage of cells during
 CC differentiation, development in pathological conditions, in tissue
 CC regeneration, in transplantation or in therapeutic and physiological
 CC repair of tissues. The pathological conditions involving subunit alpha11
 CC are selected from damage of cells, muscle dystrophy, fibrosis, wound
 CC healing, trauma, rheumatoid arthritis, osteoarthritis and osteoporosis,
 CC damage of cartilage and bone, and cartilage and bone diseases. The
 CC polypeptide is useful for detecting the formation of cartilage during
 CC embryonic development, for detecting physiological therapeutic repair of
 CC cartilage and muscle, for selection and analysis, or for sorting,
 CC isolating or purification of chondrocytes and muscle cells, for detecting
 CC regeneration of cartilage or chondrocytes during transplantation of
 CC cartilage or chondrocytes during transplantation of cartilage or
 CC chondrocytes, respectively, or of muscle or muscle cells during
 CC transplantation of muscle or muscle cells, respectively, and for studies
 CC of differentiation or chondrocytes or muscle cells
 XX
 SQ Sequence 1188 AA;
 Query Match 100.0%; Score 125; DB 4; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 8e-11;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EYQCVLRKPAQDCSAYTILSFT 23
 DB 804 EYQCVLRKPAQDCSAYTILSFT 826
 RESULT 6
 ID AAU14467
 XX AAU14467 standard; protein; 1188 AA.
 AC AAU14467;
 XX
 XX 24-OCT-2001 (first entry)
 DT
 XX

DE Human novel protein #338.
 XX
 KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 XX tissue regeneration; immune disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200155437-A2.
 XX
 XX PD 02-AUG-2001.
 XX
 XX PF 25-JAN-2001; 2001WO-US002623.
 XX
 XX PR 25-JAN-2000; 2000US-00491404.
 XX
 XX PA (HYSE-) HYSEQ INC.
 XX
 XX PI Tang YT, Liu C, Drmanac RT;
 XX
 XX DR WPI; 2001-451939/48.
 XX DR N-PSDB; AAS22772.
 XX
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage.
 PT
 XX
 XX PS Example 4; Page 828-831; 894pp; English.
 XX
 CC The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise
 CC antibodies/elicit an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention
 XX
 SQ Sequence 1188 AA;
 Query Match 100.0%; Score 125; DB 4; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 8e-11;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EYQCVLRKPAQDCSAYTILSFT 23
 DB 804 EYQCVLRKPAQDCSAYTILSFT 826
 RESULT 7
 ID AAU14231
 XX AAU14231 standard; protein; 1188 AA.
 DT
 XX

AC AAU14231;
 XX 24-OCT-2001 (first entry)
 XX Human novel protein #102.
 XX
 KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200155437-A2.
 XX 02-AUG-2001.
 XX
 XX 25-JAN-2001; 2001WO-US002623.
 XX
 XX 25-JAN-2000; 2000US-00491404.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 XX
 XX WPI; 2001-451939/48.
 XX N-PSDB; AAS22536.
 XX
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 XX nervous system disorders, and for regenerating bone and cartilage.
 XX
 XX Example 4; Page 578-581; 894pp; English.
 XX
 CC The invention relates to polynucleotides encoding novel human proteins or
 CC their active domains. The polypeptides, polynucleotides and antibodies
 CC raised against the polypeptides are used in a method of treatment of a
 CC mammal and prevention of disorders caused by the aberrant protein
 CC expression or activity. The polypeptides can be used as molecular weight
 CC markers, food supplements, and in antibody production. The polypeptides
 CC are used to identify compounds which bind to the polypeptides.
 CC Polynucleotides of the invention are used as probes and primers, for
 CC sequencing, for chromosome or gene mapping, in the production of
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene
 CC therapy. Polypeptides of the invention can be used to target drugs to a
 CC tumour, in assays to determine biological activity, to raise
 CC antibodies/elicits an immune response, to determine quantitative protein
 CC levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory
 CC diseases, nervous system disorders, and infection. The present sequence
 CC represents a protein of the invention
 XX
 XX Sequence 1188 AA;

Query Match 100.0%; Score 125; DB 4; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 8e-11;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYCORVLRKPAQDCSAYTLSPDT 23
 |||||
 Db 804 EYCORVLRKPAQDCSAYTLSPDT 826

RESULT 8
 AAB50085
 ID AAB50085 standard; protein; 1188 AA.
 XX
 XX AAB50085;
 XX
 XX 19-MAR-2001 (first entry)
 XX
 XX Human A259.
 XX
 KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
 KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
 KW rheumatoid arthritis.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Domain 1..1141
 XX /label= Extracellular_domain
 XX Peptide 1..22
 XX /label= Signal_peptide
 XX Protein 23..1188
 XX /label= Mature_protein
 XX Domain 39..74
 XX /label= Integrin_alphasubunit_repeat_domain_#1
 XX Domain 115..157
 XX /label= Integrin_alphasubunit_repeat_domain_#2
 XX Domain 164..345
 XX /label= I domain
 XX Domain 367..392
 XX /label= Integrin_alphasubunit_repeat_domain_#3
 XX Domain 421..455
 XX /label= Integrin_alphasubunit_repeat_domain_#4
 XX Domain 478..516
 XX /label= Integrin_alphasubunit_repeat_domain_#5
 XX Domain 540..575
 XX /label= Integrin_alphasubunit_repeat_domain_#6
 XX Domain 602..640
 XX /label= Integrin_alphasubunit_repeat_domain_#7
 XX Domain 1142..1164
 XX /label= Transmembrane_domain
 XX Domain 1165..1188
 XX /label= Cytoplasmic_domain
 XX
 XX WO200073339-A1.
 XX
 XX 07-DEC-2000.
 XX
 XX 15-MAY-2000; 2000WO-US013262.
 XX
 XX 28-MAY-1999; 99US-00322790.
 XX 27-APR-2000; 2000US-00561263.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 XX Pan Y, Lora JM;
 XX WPI; 2001-041142/05.
 XX N-PSDB; AAC91901, AAC91902.
 XX
 XX Nucleic acid encoding alpha-integrin subunits, useful for treatment and
 XX diagnosis of fibrosis, e.g. of the liver.
 XX
 XX Claim 8; Fig 1; 164pp; English.
 XX
 XX The present sequence is human integrin alpha subunit, A259. A259 is
 XX homologous with the alpha1 and alpha10 integrin subunits and is
 XX overexpressed in fibrosis. A259 is implicated in regulation of
 XX proliferation, differentiation and/or function of many different cell
 XX types. Inhibitors of A259 activity are useful for the treatment of liver
 XX disease, particularly fibrosis, and also fibrosis in other organs
 XX (specifically lung and kidney). In addition, A259 can be used for
 XX treatment and prevention of cancer, osteoporosis, acute myeloid

```

CC leukaemia, HIV infection, and rheumatoid arthritis
XX
SQ Sequence 1188 AA;

Query Match 100.0%; Score 125; DB 4; Length 1188;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYQQRVLRKPAQDCSAYTLSPDT 23
    |||||
Db 804 EYQQRVLRKPAQDCSAYTLSPDT 826

RESULT 9
AAU10551
ID AAU10551 standard; protein; 1188 AA.
AC AAU10551;
XX
DT 14-FEB-2002 (first entry)
DE Human A259 polypeptide.
XX
KW Human; A259; integrin alpha subunit; integrin alpha 10; secreted protein;
KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
KW cartilage associated disorder; haematopoietic disorder; bone marrow;
KW immune related disease; apoptotic disorder; neuronal tissue disease;
KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
KW antiarthritic; antianaemic; antiallergic; antiasthmatic; dermatological;
KW antidiabetic; anticonvulsant; antiparkinsonian.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1..1141
FT /note= "Extracellular domain"
FT Peptide 1...22
FT /note= "Signal peptide"
FT Protein 23..1188
FT /note= "Mature human A259"
FT Domain 37..90
FT /note= "Integrin alpha repeat domain. The specification
FT states that this domain exists in human A549"
FT Domain 115..117
FT /note= "Integrin alpha repeat domain"
FT Domain 164..345
FT /note= "I domain or Von Willebrand Factor type A domain"
FT Domain 367..392
FT /note= "Integrin alpha repeat domain"
FT Domain 421..472
FT /note= "Integrin alpha repeat domain. The specification
FT states that this domain exists in human A549"
FT Domain 476..532
FT /note= "Integrin alpha repeat domain. The specification
FT states that this domain exists in human A549"
FT Domain 538..593
FT /note= "Integrin alpha repeat domain. The specification
FT states that this domain exists in human A549"
FT Domain 600..654
FT /note= "Integrin alpha repeat domain. The specification
FT states that this domain exists in human A549"
FT Domain 1142..1164
FT /note= "Transmembrane domain"
FT Domain 1165..1188
FT /note= "Cytoplasmic domain"
XX
WO200181414-A2.
XX
PD 01-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US013516.
XX

```

```

PR 27-APR-2000; 2000US-00561263.
XX (MILL-) MILLENNIUM PHARM INC.
XX Pan Y, Lora J;
XX WPI; 2002-041397/05.
DR N-PSDB; AAS16873.
DR
XX New A259 nucleic acids and polypeptides, which comprise integrin alpha
PT subunit, useful for diagnosing, preventing or treating e.g. liver
PT disease, kidney or lung fibrosis, cancers, blood disorders or immune
PT related diseases.
XX
PS Claim 9; Fig 1; 168pp; English.
XX
CC The invention relates to human and murine A259 nucleic acid molecules
CC which encode secreted proteins with homology to integrin alpha subunits,
CC specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
CC are useful for treating liver disease or fibrosis, particularly kidney
CC fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
CC useful for diagnosing, preventing or treating cartilage and bone
CC associated disorders (such as bone cancer, achondroplasia, myeloma,
CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
CC osteoporosis), bone marrow, blood and haematopoietic disorders (such as
CC acute myeloid leukaemia, haemophilia, anaemia and thalassaemia), immune
CC related diseases (such as HIV, viral infections, cancers, T cell
CC autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
CC asthma and psoriasis), apoptotic disorders (such as systemic lupus
CC erythematosus and insulin-dependent diabetes mellitus), diseases of the
CC neuronal tissues (such as epilepsy and muscular dystrophy) and
CC neurodegenerative diseases (such as Parkinson's disease and Huntington's
CC disease). This sequence represents the human A259 polypeptide
XX
SQ Sequence 1188 AA;

Query Match 100.0%; Score 125; DB 5; Length 1188;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYQQRVLRKPAQDCSAYTLSPDT 23
    |||||
Db 804 EYQQRVLRKPAQDCSAYTLSPDT 826

RESULT 10
ADE09956
ID ADE09956 standard; protein; 1188 AA.
XX
AC ADE09956;
XX
DT 29-JAN-2004 (first entry)
DE Novel protein-related contig polypeptide sequence #544.
XX
KW novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder; contig.
XX
OS Unidentified.
XX
PN WO2003054152-A2.
XX
PD 03-JUL-2003.
XX
PF 10-DEC-2002; 2002WO-US039555.
XX
PR 10-DEC-2001; 2001US-0339739P.
PR 11-DEC-2001; 2001US-0339453P.
PR 14-MAR-2002; 2002US-0365091P.
PR 14-MAR-2002; 2002US-0365384P.
PR 12-APR-2002; 2002US-0372381P.
PR 12-APR-2002; 2002US-0372615P.
PR 22-APR-2002; 2002US-00128558.

```


PR 24-APR-2002; 2002US-0376045P.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX
XX WPI; 2003-569235/53.
XX
XX New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX
XX
PS Disclosure; SEQ ID NO 3022; 1177pp; English.
XX
XX The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence was used in the
XX exemplification of the invention.
XX
SQ Sequence 1188 AA;

Query Match 100.0%; Score 125; DB 7; Length 1188;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYCORVLRKPAQDCSAYTSLSFDT 23
DB 804 EYCORVLRKPAQDCSAYTSLSFDT 826
|||||
RESULT 11
ID AAB25582
XX AAB25582 standard; protein; 1189 AA.
XX
AC AAB25582;
XX
XX 21-NOV-2000 (first entry)
XX
XX ITGAl1 protein encoded by human secreted protein gene #7.
XX
XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
KW antirheumatic, dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnery; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
XX
XX Homo sapiens.
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX 27-OCT-1999; 99WO-US025031.
XX
XX 28-OCT-1998; 98US-0105971P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
PI Greene JM;
XX
XX WPI; 2000-387742/33.
XX
XX N-PSDB; AAA80612.
XX

PT Isolated nucleic acid molecules encoding human secreted proteins are used
PT for the prevention, amelioration and treatment of autoimmune,
PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
PT wounds, and infectious diseases.
XX
XX Claim 1; Fig 19A-F; 803pp; English.
XX
XX The present invention relates to 12 secreted human proteins and the
CC nucleotide sequences encoding them. The polynucleotide sequences given in
CC AAB80606-A80623 encode the 12 secreted protein sequences given in
CC AAB25576-B25593. The human secreted proteins have various activities
CC dependent on the tissues in which they are expressed. Examples of the
CC activities of the proteins include: immunosuppressant; anti-inflammatory;
CC antiarthritic; antirheumatic, dermatological; antiproliferative;
CC antiarteriosclerotic; anticancer; vulnery; antiviral; antibacterial;
CC and antifungal activity. The proteins, polypeptides, agonists and
CC antagonists may be used to treat prevent and/or diagnose various disease,
CC disorders and conditions examples of which include: immune disorders e.g.
CC Addison's disease, rheumatoid arthritis, dermatitis, and multiple
CC sclerosis; inflammatory disorders e.g. inflammatory bowel disease,
CC Crohn's disease and nephritis; hyperproliferative disorders such as
CC paraproteinemias and purpura; cardiovascular disorders e.g. coronary
CC arteriosclerosis and myocarditis; cancer e.g. melanoma and lymphoma. The
CC proteins and polynucleotide sequences may also be used in wound healing
CC and the treatment of infectious diseases. The human secreted protein gene
CC #7 and protein sequences are represented in sequences AAA80612 and
CC AAB25582. Secreted protein gene #7 is located at position chromosome 15
CC q22.3-23. Sequences AAA80652-A80661 represent genes which are related to
CC the secreted protein gene#7
XX
XX Sequence 1189 AA;

Query Match 100.0%; Score 125; DB 3; Length 1189;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYCORVLRKPAQDCSAYTSLSFDT 23
DB 804 EYCORVLRKPAQDCSAYTSLSFDT 826
|||||
RESULT 12
ID AAB25582
XX AAB25582 standard; protein; 1189 AA.
XX
AC AAB25582;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #12940.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS77136.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in

Query Match 100.0%; Score 125; DB 7; Length 1189;
 Best Local Similarity 100.0%; Pred. No. 8e-11;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYQQRVLRKPAQDCSAYTLSPDT 23
 |||||
 Db 804 EYQQRVLRKPAQDCSAYTLSPDT 826

Search completed: September 21, 2004, 12:59:15
 Job time : 6.38478 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 13:03:59 : Search time 4.14395 Seconds
(without alignments)
1782.414 Million cell updates/sec

Title: US-09-980-403-2_COPY_804_826

Perfect score: 125

Sequence: 1 EYQCVLRKPAQDCSAYTSLSFD 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125	100.0	1034	10	US-09-984-130-43
2	125	100.0	1034	10	US-09-836-353A-43
3	125	100.0	1120	12	US-10-262-839-6
4	125	100.0	1188	15	US-10-291-265-338
5	125	100.0	1188	15	US-10-291-265-810
6	125	100.0	1189	10	US-09-984-130-35
7	125	100.0	1189	10	US-09-836-353A-35
8	125	100.0	1189	12	US-10-262-839-4
9	115	92.0	545	10	US-09-866-050A-500
10	115	92.0	688	10	US-09-866-050A-624
11	115	92.0	696	10	US-09-866-050A-501
12	51.5	41.2	897	16	US-10-437-963-113954
13	48	38.4	1184	9	US-09-757-716-3
14	48	38.4	1184	9	US-09-989-722-124
15	48	38.4	1184	9	US-09-989-723-124

16	48	38.4	1184	9	US-09-989-279-124	Sequence 124, App
17	48	38.4	1184	9	US-09-989-727-124	Sequence 124, App
18	48	38.4	1184	9	US-09-989-731-124	Sequence 124, App
19	48	38.4	1184	9	US-09-989-732-124	Sequence 124, App
20	48	38.4	1184	9	US-09-991-073-124	Sequence 124, App
21	48	38.4	1184	9	US-09-990-442-124	Sequence 124, App
22	48	38.4	1184	9	US-09-991-163-124	Sequence 124, App
23	48	38.4	1184	9	US-09-993-604-124	Sequence 124, App
24	48	38.4	1184	9	US-09-990-456-124	Sequence 124, App
25	48	38.4	1184	9	US-09-989-721-124	Sequence 124, App
26	48	38.4	1184	9	US-09-992-598-124	Sequence 124, App
27	48	38.4	1184	9	US-09-989-293A-124	Sequence 124, App
28	48	38.4	1184	9	US-09-989-735-124	Sequence 124, App
29	48	38.4	1184	9	US-09-990-444-124	Sequence 124, App
30	48	38.4	1184	9	US-09-991-181-124	Sequence 124, App
31	48	38.4	1184	9	US-09-989-730-124	Sequence 124, App
32	48	38.4	1184	9	US-09-990-436-124	Sequence 124, App
33	48	38.4	1184	9	US-09-993-687-124	Sequence 124, App
34	48	38.4	1184	10	US-09-989-734-124	Sequence 124, App
35	48	38.4	1184	10	US-09-997-653-124	Sequence 124, App
36	48	38.4	1184	10	US-09-993-667-124	Sequence 124, App
37	48	38.4	1184	10	US-09-997-428-124	Sequence 124, App
38	48	38.4	1184	10	US-09-997-666-124	Sequence 124, App
39	48	38.4	1184	10	US-09-990-438-124	Sequence 124, App
40	48	38.4	1184	10	US-09-990-562-124	Sequence 124, App
41	48	38.4	1184	10	US-09-990-711-124	Sequence 124, App
42	48	38.4	1184	10	US-09-989-726-124	Sequence 124, App
43	48	38.4	1184	10	US-09-998-156-124	Sequence 124, App
44	48	38.4	1184	10	US-09-990-437-124	Sequence 124, App
45	48	38.4	1184	10	US-09-991-157-124	Sequence 124, App

ALIGNMENTS

RESULT 1

US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43

Query Match 100.0%; Score 125; DB 10; Length 1034;
Best Local Similarity 100.0%; Pred. No. 9.9e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQCVLRKPAQDCSAYTSLSFD 23

Db 804 EYQCVLRKPAQDCSAYTSLSFD 826

RESULT 2

US-09-836-353A-43
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-43

Query Match 100.0%; Score 125; DB 10; Length 1034;
Best Local Similarity 100.0%; Pred. No. 9.9e-11;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQCVLRKPAQDCSAYTLSEDT 23
|||||
DB 804 EYQCVLRKPAQDCSAYTLSEDT 826

RESULT 3

US-10-262-839-6
; Sequence 6, Application US/10262839
; Publication No. US2004003887A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Holdog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Ellina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 6
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-6

Query Match 100.0%; Score 125; DB 12; Length 1120;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQCVLRKPAQDCSAYTLSEDT 23
|||||
DB 735 EYQCVLRKPAQDCSAYTLSEDT 757

RESULT 4

US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338

Query Match 100.0%; Score 125; DB 15; Length 1188;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQCVLRKPAQDCSAYTLSEDT 23
|||||
DB 804 EYQCVLRKPAQDCSAYTLSEDT 826

RESULT 5

US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: PastSeq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-810

Query Match 100.0%; Score 125; DB 15; Length 1189;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYCORVLRKPAQDCSAYTLSDT 23
|||||
Db 804 EYCORVLRKPAQDCSAYTLSDT 826

RESULT 6
US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US2003005231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489p2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-35

Query Match 100.0%; Score 125; DB 10; Length 1189;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYCORVLRKPAQDCSAYTLSDT 23
|||||
Db 804 EYCORVLRKPAQDCSAYTLSDT 826

RESULT 7
US-09-836-353A-35
; Sequence 35, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:

; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489p1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-35

Query Match 100.0%; Score 125; DB 10; Length 1189;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EYCORVLRKPAQDCSAYTLSDT 23
|||||
Db 804 EYCORVLRKPAQDCSAYTLSDT 826

RESULT 8
US-10-262-839-4
; Sequence 4, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16

```
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-4

Query Match          100.0%; Score 125; DB 12; Length 1189;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQQRVLRKPAQDCSAYTSLSFT 23
Db 804 EYQQRVLRKPAQDCSAYTSLSFT 826

RESULT 9
US-09-866-050A-500
; Sequence 500, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 500
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-500

Query Match          92.0%; Score 115; DB 10; Length 545;
Best Local Similarity 87.0%; Pred. No. 1.9e-09;
Matches 20; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQQRVLRKPAQDCSAYTSLSFT 23
Db 161 EYQQRVLRKPAQDCSAYTSLSFT 183

RESULT 10
US-09-866-050A-624
; Sequence 624, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
```

```
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-624

Query Match          92.0%; Score 115; DB 10; Length 688;
Best Local Similarity 87.0%; Pred. No. 2.4e-09;
Matches 20; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQQRVLRKPAQDCSAYTSLSFT 23
Db 304 EYQQRVLRKPAQDCSAYTSLSFT 326

RESULT 11
US-09-866-050A-501
; Sequence 501, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-501

Query Match          92.0%; Score 115; DB 10; Length 696;
Best Local Similarity 87.0%; Pred. No. 2.4e-09;
Matches 20; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EYQQRVLRKPAQDCSAYTSLSFT 23
Db 304 EYQQRVLRKPAQDCSAYTSLSFT 326

RESULT 12
US-10-437-963-113954
; Sequence 113954, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```



```

; SEQ ID NO 113954
; LENGTH: 897
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17693C.1.pep
US-10-437-963-113954

Query Match
Best Local Similarity 41.2%; Score 51.5; DB 16; Length 897;
Matches 12; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

QY 1 EYCORVLRKPAQDCSAYTLSP 21
||| ||| ||| ||| ||| |||
Db 601 EYCEPTLR--QDPETYISF 618

RESULT 13
US-09-757-716-3
; Sequence 3, Application US/09757716
; Patent No. US20010012515A1
; GENERAL INFORMATION:
; APPLICANT: Magna, Holly
; APPLICANT: Schaffer, Paul
; APPLICANT: Lawton, Michael
; APPLICANT: Yocum, Sue
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHODIOLASE-2
; FILE REFERENCE: PP-0420 US
; CURRENT APPLICATION NUMBER: US/09/757,716
; CURRENT FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: US/08/996,083
; PRIOR FILING DATE: 1997-12-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone No. US20010012515A1: 422069
; PUBLICATION INFORMATION:
US-09-757-716-3

Query Match
Best Local Similarity 38.4%; Score 48; DB 9; Length 1184;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 YCORVLRKPAQDCSAYTLSP 21
||| ||| ||| ||| ||| |||
Db 118 WCLNRQRPQNCNNTVRF 137

RESULT 14
US-09-989-722-124
; Sequence 124, Application US/09989722
; Patent No. US20020072067A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher

```

; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089440
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089600
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089908
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089948
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/089952
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090246
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090252
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090254
; PRIOR FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540

; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 38.4%; Score 48; DB 9; Length 1184;

Best Local Similarity 40.0%; Pred. No. 1.5e+02;

Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 YCQVLRKPAQDCSAYTLSE 21

Db 118 WCLNREQPGQNCNVTURF 137

RESULT 15

US-09-989-723-124
; Sequence 124, Application US/09889723
; Patent No. US20020072092A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Borstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerbitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P27301C62
CURRENT APPLICATION NUMBER: US/09/989,723
PRIORITY FILING DATE: 2001-11-19
PRIORITY APPLICATION NUMBER: 60/049787
PRIORITY FILING DATE: 1997-06-16
PRIORITY APPLICATION NUMBER: 60/062250
PRIORITY FILING DATE: 1997-10-17
PRIORITY APPLICATION NUMBER: 60/065186
PRIORITY FILING DATE: 1997-11-12
PRIORITY APPLICATION NUMBER: 60/065311
PRIORITY FILING DATE: 1997-11-13
PRIORITY APPLICATION NUMBER: 60/066770
PRIORITY FILING DATE: 1997-11-24
PRIORITY APPLICATION NUMBER: 60/075945
PRIORITY FILING DATE: 1998-02-25
PRIORITY APPLICATION NUMBER: 60/078910
PRIORITY FILING DATE: 1998-03-20
PRIORITY APPLICATION NUMBER: 60/083322
PRIORITY FILING DATE: 1998-04-28
PRIORITY APPLICATION NUMBER: 60/084600
PRIORITY FILING DATE: 1998-05-07
PRIORITY APPLICATION NUMBER: 60/087106
PRIORITY FILING DATE: 1998-05-28
PRIORITY APPLICATION NUMBER: 60/087607
PRIORITY FILING DATE: 1998-06-02
PRIORITY APPLICATION NUMBER: 60/087609
PRIORITY FILING DATE: 1998-06-02
PRIORITY APPLICATION NUMBER: 60/087759
PRIORITY FILING DATE: 1998-06-02
PRIORITY APPLICATION NUMBER: 60/087827
PRIORITY FILING DATE: 1998-06-03
PRIORITY APPLICATION NUMBER: 60/088021
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088025
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088026
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088028
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088029
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088030
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088033
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088326
PRIORITY FILING DATE: 1998-06-04
PRIORITY APPLICATION NUMBER: 60/088167
PRIORITY FILING DATE: 1998-06-05
PRIORITY APPLICATION NUMBER: 60/088202
PRIORITY FILING DATE: 1998-06-05
PRIORITY APPLICATION NUMBER: 60/088212
PRIORITY FILING DATE: 1998-06-05
PRIORITY APPLICATION NUMBER: 60/088217
PRIORITY FILING DATE: 1998-06-05
PRIORITY APPLICATION NUMBER: 60/088655
PRIORITY FILING DATE: 1998-06-09
PRIORITY APPLICATION NUMBER: 60/088734
PRIORITY FILING DATE: 1998-06-10
PRIORITY APPLICATION NUMBER: 60/088738
PRIORITY FILING DATE: 1998-06-10
PRIORITY APPLICATION NUMBER: 60/088742
PRIORITY FILING DATE: 1998-06-10
PRIORITY APPLICATION NUMBER: 60/088810
PRIORITY FILING DATE: 1998-06-10
PRIORITY APPLICATION NUMBER: 60/088824
PRIORITY FILING DATE: 1998-06-10
PRIORITY APPLICATION NUMBER: 60/088826
PRIORITY FILING DATE: 1998-06-10
PRIORITY APPLICATION NUMBER: 60/088858
PRIORITY FILING DATE: 1998-06-11
PRIORITY APPLICATION NUMBER: 60/088861
PRIORITY FILING DATE: 1998-06-11
PRIORITY APPLICATION NUMBER: 60/088876
PRIORITY FILING DATE: 1998-06-11
PRIORITY APPLICATION NUMBER: 60/089105
PRIORITY FILING DATE: 1998-06-12
PRIORITY APPLICATION NUMBER: 60/089440
PRIORITY FILING DATE: 1998-06-16
PRIORITY APPLICATION NUMBER: 60/089512
PRIORITY FILING DATE: 1998-06-16
PRIORITY APPLICATION NUMBER: 60/089514
PRIORITY FILING DATE: 1998-06-16
PRIORITY APPLICATION NUMBER: 60/089532
PRIORITY FILING DATE: 1998-06-17
PRIORITY APPLICATION NUMBER: 60/089538
PRIORITY FILING DATE: 1998-06-17
PRIORITY APPLICATION NUMBER: 60/089598
PRIORITY FILING DATE: 1998-06-17
PRIORITY APPLICATION NUMBER: 60/089599
PRIORITY FILING DATE: 1998-06-17
PRIORITY APPLICATION NUMBER: 60/089600
PRIORITY FILING DATE: 1998-06-17
PRIORITY APPLICATION NUMBER: 60/089653
PRIORITY FILING DATE: 1998-06-17
PRIORITY APPLICATION NUMBER: 60/089801
PRIORITY FILING DATE: 1998-06-18
PRIORITY APPLICATION NUMBER: 60/089907
PRIORITY FILING DATE: 1998-06-18
PRIORITY APPLICATION NUMBER: 60/089908
PRIORITY FILING DATE: 1998-06-18
PRIORITY APPLICATION NUMBER: 60/089947
PRIORITY FILING DATE: 1998-06-19
PRIORITY APPLICATION NUMBER: 60/089948
PRIORITY FILING DATE: 1998-06-19
PRIORITY APPLICATION NUMBER: 60/089952
PRIORITY FILING DATE: 1998-06-19
PRIORITY APPLICATION NUMBER: 60/090246
PRIORITY FILING DATE: 1998-06-22
PRIORITY APPLICATION NUMBER: 60/090252
PRIORITY FILING DATE: 1998-06-22
PRIORITY APPLICATION NUMBER: 60/090254
PRIORITY FILING DATE: 1998-06-22
PRIORITY APPLICATION NUMBER: 60/090349
PRIORITY FILING DATE: 1998-06-23
PRIORITY APPLICATION NUMBER: 60/090355
PRIORITY FILING DATE: 1998-06-23
PRIORITY APPLICATION NUMBER: 60/090429
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090431
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090435
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090444
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090445
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090472
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090535
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090540
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090542
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090557
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090676
PRIORITY FILING DATE: 1998-06-25

```

; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

Query Match 38.4%; Score 48; DB 9; Length 1184;

Best Local Similarity 40.0%; Pred. No. 1.5e+02;

Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 2 YCQVLRKPAQDCSAYTLSE 21

DB 118 WCLNREQPGQNCNVTYRF 137

Search completed: September 21, 2004, 13:29:03
Job time : 5.14385 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:54:33 ; Search time 1.22067 Seconds
(without alignments)
972.742 Million cell updates/sec

Title: US-09-980-403-2_COPY_804_826

Perfect score: 125

Sequence: 1 EYQQRVLRKPAQDCSAVTLSDFT 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	38.4	1184	2	US-08-918-914-1
2	48	38.4	1184	3	US-08-996-083-3
3	46.5	37.2	1436	2	US-08-652-971-2
4	46.5	37.2	1436	2	US-08-991-258A-2
5	46.5	37.2	1436	2	US-08-769-399-2
6	46.5	37.2	1436	3	US-08-991-953A-2
7	45	36.0	566	1	US-08-073-383-4
8	45	36.0	566	1	US-08-428-415-4
9	45	36.0	566	1	US-08-379-685-4
10	45	36.0	566	2	US-08-854-029-4
11	45	36.0	566	3	US-08-848-810-2
12	45	36.0	566	3	US-08-428-762-4
13	45	36.0	566	5	PCT-US94-06365-4
14	43	34.4	825	4	US-09-489-039A-8469
15	43	34.4	924	2	US-08-588-983-18
16	43	34.4	924	2	US-08-588-976-18
17	42	33.6	162	4	US-09-252-991A-16995
18	42	33.6	235	4	US-09-029-785-1
19	42	33.6	849	4	US-09-081-385-152
20	42	33.6	2491	4	US-09-207-363-1
21	41.5	33.2	355	4	US-09-489-039A-10767
22	41	32.8	170	6	5215917-2
23	41	32.8	170	6	5472844-2
24	41	32.8	395	4	US-09-328-352-7437
25	41	32.8	1056	4	US-09-134-000C-5086
26	40.5	32.4	209	4	US-09-252-991A-20295
27	40.5	32.4	328	4	US-09-489-039A-7290

Sequence 6, Appli
Sequence 6, Appli
Sequence 6, Appli
Sequence 8, Appli
Sequence 256, App
Sequence 6, Appli
Sequence 22424, A
Sequence 14017, A
Sequence 25, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 30291, A
Sequence 18554, A
Patent No. 5171685

ALIGNMENTS

RESULT 1

US-08-918-914-1
; Sequence 1, Application US/08918914
; Patent No. 5878963
; GENERAL INFORMATION:
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Lawton, Michael
; APPLICANT: Magna, Holly
; APPLICANT: Yocum, Sue
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/918,914
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1184 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: ???
; CLONE: 422069
US-08-918-914-1

Query Match 38.4%; Score 48; DB 2; Length 1184;
Best Local Similarity 40.0%; Pred. No. 40;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 2 YCQVLRKPAQDCSAYTILSF 21
Db 118 WCLNREQRPGQNCNYYTVRF 137

RESULT 2
US-08-996-083-3
; Sequence 3, Application US/08996083A
; Patent No. 6124095
; GENERAL INFORMATION:
; APPLICANT: Magna, Holly
; APPLICANT: Schaffer, Paul
; APPLICANT: Lawton, Michael
; APPLICANT: Yocum, Sue
; APPLICANT: Mitchell, Peter
; APPLICANT: Hutchinson, Nancy
; APPLICANT: Murty, Lynn E.
; TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
; FILE REFERENCE: PF-0420 US
; CURRENT APPLICATION NUMBER: US/08/996,083A
; CURRENT FILING DATE: 1997-12-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone No. 6124095: 422069
; PUBLICATION INFORMATION:
US-08-996-083-3

Query Match 38.4%; Score 48; DB 3; Length 1184;
Best Local Similarity 40.0%; Pred. No. 40;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 2 YCQVLRKPAQDCSAYTILSF 21
Db 118 WCLNREQRPGQNCNYYTVRF 137

RESULT 3
US-08-652-971-2
; Sequence 2, Application US/08652971
; Patent No. 5814507
; GENERAL INFORMATION:
; APPLICANT: Cheng, Jill
; APPLICANT: Lasky, Laurence A.
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd.
; CITY: South San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,971
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 225-3216
; TELEFAX: (415) 952-9881
; TELEX: 910 371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-652-971-2

Query Match 37.2%; Score 46.5; DB 2; Length 1436;
Best Local Similarity 47.6%; Pred. No. 85;
Matches 10; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Qy 4 QRVLRKP-AQDCSAYTILSFDT 23
Db 635 RRLRREGAQDCFSVPLTFET 655

RESULT 4
US-08-991-258A-2
; Sequence 2, Application US/08991258A
; Patent No. 5928887
; GENERAL INFORMATION:
; APPLICANT: Cheng, Jill
; APPLICANT: Lasky, Laurence A.
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT, LLP
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,258A
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/652,971
; FILING DATE: 24-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Walter H.
; REGISTRATION NUMBER: 24,190
; REFERENCE/DOCKET NUMBER: A-63478-3/WHD/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-991-258A-2

Query Match 37.2%; Score 46.5; DB 2; Length 1436;
Best Local Similarity 47.6%; Pred. No. 85;
Matches 10; Conservative 6; Mismatches 4; Indels 1; Gaps 1;


```

; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-073-383-4

```

```

Query Match 36.0%; Score 45; DB 1; Length 566;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 CORVLKPAQDCS 15
DB 301 CORLFRSPSMPCS 313

```

RESULT 8

```

US-08-428-415-4
; Sequence 4, Application US/08428415
; Patent No. 5756335
; GENERAL INFORMATION:
; APPLICANT: Cold Spring Harbor Laboratory
; TITLE OF INVENTION: No. 5756335el Human cdc25 Genes, Encoded
; TITLE OF INVENTION: Products and Uses Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,415
; FILING DATE: 24 April 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-019CP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-428-415-4

```

```

Query Match 36.0%; Score 45; DB 1; Length 566;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 CORVLKPAQDCS 15
DB 301 CORLFRSPSMPCS 313

```

RESULT 9

```

US-08-379-685-4
; Sequence 4, Application US/08379685
; Patent No. 5770423
; GENERAL INFORMATION:
; APPLICANT: Cold Spring Harbor Laboratory

```

```

; TITLE OF INVENTION: No. 5770423el Human cdc25 Genes, Encoded
; TITLE OF INVENTION: Products and Uses Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/379,685
; FILING DATE: 26 January 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-019-DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-379-685-4

```

```

Query Match 36.0%; Score 45; DB 1; Length 566;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 CORVLKPAQDCS 15
DB 301 CORLFRSPSMPCS 313

```

RESULT 10

```

US-08-854-029-4
; Sequence 4, Application US/08854029
; Patent No. 5994074
; GENERAL INFORMATION:
; APPLICANT: Cold Spring Harbor Laboratory
; TITLE OF INVENTION: No. 5994074el Human cdc25 Genes, Encoded
; TITLE OF INVENTION: Products and Uses Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,029
; FILING DATE: 2 MAY 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MIV-019.08
; TELECOMMUNICATION INFORMATION:

```



```
; TELEPHONE: 617-832-1242
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-854-029-4

Query Match 36.0%; Score 45; DB 2; Length 566;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 CORVLKPAQDCS 15
Db 301 CQRLFRSPMPCS 313

RESULT 11
US-08-848-810-2
; Sequence 2, Application US/08848810
; Patent No. 6074851
; GENERAL INFORMATION:
; APPLICANT: Deibel Jr., M. R.
; APPLICANT: Yem, A. W.
; APPLICANT: Wilson, C. L.
; TITLE OF INVENTION: Catalytic Macro Molecules Having DCD25B
; TITLE OF INVENTION: Like Activity
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: MI
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/848,810
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wootton, Thomas A.
; REGISTRATION NUMBER: 35,004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-7914
; TELEFAX: 616-833-6897
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-848-810-2

Query Match 36.0%; Score 45; DB 3; Length 566;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 CORVLKPAQDCS 15
Db 301 CQRLFRSPMPCS 313

RESULT 12
```

```
US-08-428-762-4
; Sequence 4, Application US/08428762
; Patent No. 6251585
; GENERAL INFORMATION:
; APPLICANT: Cold Spring Harbor Laboratory
; TITLE OF INVENTION: No. 6251585el Human cdc25 Genes, Encoded
; TITLE OF INVENTION: Products and Uses Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,762
; FILING DATE: 24 April 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-019-DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-227-7400
; TELEFAX: 617-227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-428-762-4

Query Match 36.0%; Score 45; DB 3; Length 566;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 CORVLKPAQDCS 15
Db 301 CQRLFRSPMPCS 313

RESULT 13
PCT-US94-06365-4
; Sequence 4, Application PC/TUS9406365
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay and Reagents for Identifying
; TITLE OF INVENTION: Anti-proliferative Agents
; NUMBER OF SEQUENCES: 6
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06365
; FILING DATE: 06-JUN-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,383
; FILING DATE: 04-JUN-1993
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 566 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

; MOLECULE TYPE: protein
PCT-US94-06365-4

Query Match 36.0%; Score 45; DB 5; Length 566;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 CORVLKPAQDCS 15
|||: |||: |||
Db 301 CORLFRSPMPCS 313

RESULT 14
US-09-489-039A-8469
; Sequence 8469, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8469
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8469

Query Match 34.4%; Score 43; DB 4; Length 825;
Best Local Similarity 64.3%; Pred. No. 1.7e+02;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 10 PAQDCSAYTLSPDT 23
Db 624 PGSDNSYYKLSPDT 637

RESULT 15
US-08-588-983-18
; Sequence 18, Application US/08588983
; Patent No. 5854067
; GENERAL INFORMATION:
; APPLICANT: Christopher B. Newgard, et al.
; TITLE OF INVENTION: Methods and Compositions
; TITLE OF INVENTION: for Inhibiting Hexokinase
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,983
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: UTSD:424/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577

; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 924 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-588-983-18

Query Match 34.4%; Score 43; DB 2; Length 924;
Best Local Similarity 56.2%; Pred. No. 1.9e+02;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 BYCORVLRKPAQDCSA 16
|||: |||: |||: |||: |||: |||
Db 828 EVCQAVSRRRAQLCGA 843

Search completed: September 21, 2004, 13:06:21
Job time : 2.22067 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:52:17 ; Search time 9.90503 Seconds
(without alignments)
1913.143 Million cell updates/sec

Title: US-09-980-403-2_COPY_159_355

Perfect score: 1022

Sequence: 1 CQTYMDIVLDGNSIYPW.....AALKDIVDAGDRIFSLEGT 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	599	58.6	1151	2 A45226	integrin alpha-1 c
2	599	58.6	1180	2 A35854	integrin alpha-1 c
3	598	58.5	272	2 A55348	integrin alpha-1 -
4	486	47.6	1178	2 S44142	VLA-2 protein homo
5	468	45.8	1170	2 I45914	integrin alpha 2 s
6	465	45.5	1181	2 A33998	integrin alpha-2 c
7	408	39.9	191	2 I47230	VLA-2 protein - pi
8	248.5	24.3	3124	2 A40020	collagen alpha 1(X
9	244.5	23.9	1179	2 A53213	integrin alpha-E c
10	240.5	23.5	1153	2 S00551	leukocyte surface
11	237.5	23.2	929	2 I51027	type XII collagen
12	235.5	23.0	1747	2 A45974	collagen alpha 1(X
13	234.5	22.9	1857	2 S31212	collagen alpha 1(X
14	234.5	22.9	1888	2 S78476	collagen alpha 1(X
15	223.5	21.9	1163	1 RWHUIC	cell surface glyco
16	216.5	21.2	1153	1 RWHUIC	cell surface glyco
17	210	20.5	741	2 T46488	hypothetical prote
18	199.5	19.5	3137	2 A37797	collagen alpha 3(V
19	194.5	19.0	496	2 A37979	collagen matrix p
20	193	18.9	500	2 S66522	cartilage matrix p
21	193	18.9	396	2 GGH3A	cartilage matrix p
22	189.5	18.5	493	2 A33809	collagen alpha 3(V
23	188.5	18.4	843	2 A40970	cartilage matrix p
24	188	18.4	2944	2 A54849	undulin 1 - human
25	176	17.2	3051	2 S42373	collagen alpha 1(V
26	173	16.9	1170	2 S03308	hypothetical prote
27	157	15.4	341	2 T32949	cell surface glyco
28	150.5	14.7	724	2 A48569	hypothetical prote
29	148.5	14.5	1163	2 I56126	lymphocyte function

```

30 142.5 13.9 13055 2 T16580 hypothetical prote
31 141 13.8 427 2 G00039 von Willebrand fac
32 132.5 13.0 2813 1 VWHU von Willebrand fac
33 131 12.8 550 2 T23760 hypothetical prote
34 126.5 12.4 1029 1 S21369 collagen alpha 2(V
35 126.5 12.4 2098 2 T18397 protein CTRP - mal
36 125.5 12.3 780 2 A34102 von Willebrand fac
37 122 11.9 414 2 F80323 von Willebrand fac
38 119 11.6 712 2 A45638 immunodominant mic
39 118 11.5 643 2 T19549 hypothetical prote
40 117.5 11.5 238 2 C35243 collagen alpha 2(V
41 117.5 11.5 917 2 S09646 collagen alpha 2(V
42 117.5 11.5 918 2 S23377 collagen alpha 2(V
43 117.5 11.5 1018 1 GGHU2A collagen alpha 2(V
44 117.5 11.5 1022 2 S04111 collagen alpha 2(V
45 116.5 11.4 661 2 T16597 hypothetical prote

```

ALIGNMENTS

```

RESULT 1
A45226
integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 15-Sep-2003
C:Accession: A45226
R:Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Experimental source: hepatoblastoma cell line HepG2
A>Note: sequence extracted from NCBI backbone (NCBI:P:124326)
F:142-317/Domain: von Willebrand factor type A repeat homology <VWAL>

```

```

Query Match 58.6%; Score 599; DB 2; Length 1151;
Best Local Similarity 58.3%; Pred. No. 1.6e-41;
Matches 116; Conservative 33; Mismatches 48; Indels 2; Gaps 1;

QY 1 CQTYMDIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIOGVVOYGEDVHFEFLND 60
DB 139 CSTQLDLIVLDGNSIYPWDSVTFALNDLLKMDIGPKQTQGVIGYGENVTHEFLNK 198
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGS 118
DB 199 YSSTEEVLVAKKIVQGGGQTWTALGTDARKEATEARGARGVKVMVITDGS 258
QY 119 SPLEKVIQOSRDNTRYAVAVLYGYNRRGINPFTLNEIKVIASDPDKHFNVTDEA 178
DB 259 NHRLLKKVIQDCEDENIQRFSAILGSYNRGNLSTKPFVEIKIASEPTEKHFNFV 318
QY 179 ALKXDIVDALGDRIFSLEGT 197
DB 319 ALVTIVTKLGERIFALEAT 337

```

```

RESULT 2
A35854
integrin alpha-1 chain precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 15-Sep-2003
C:Accession: A35854; S11243
R:Ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto
J. Cell Biol. 111, 709-720, 1990
A:Title: Molecular cloning of the rat integrin alpha-1-subunit: a receptor for laminin
A:Reference number: A35854; MUID:90338125; PMID:2380249
A:Accession: A35854
A>Status: preliminary
A:Molecule type: mRNA

```

A;Residues: 1-1180 <IGN>
A;Cross-references: GB:X52140; NID:g56493; PIDN:CAA36384.1.; PID:g56494
C;Keywords: cell adhesion; cytoskeleton; transmembrane protein
F;170-345/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 58.6%; Score 599; DB 2; Length 1180;
Best Local Similarity 57.8%; Pred. No. 1.6e-41;
Matches 115; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 CQTMDIVIVLDGNSNIYPWVEVQHFLINILKKFYIGPGQIQGVVGVEDVVHFEHLND 60
DB :|||:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 167 CSTQLDIIIVLDGNSNIYPWESVIAPLNDLLKEMDIGPKQTQVGIVQGENVTHEFNLNK 256

QY 61 YRSVKDVWEARASHIEQRGGTETRTAGIETPAREAF--QKGGRKGAKKMIVITDGESH 118
DB :||:||||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 227 YSSTEVELVAANKIGRGGGTCTMTALGIDTARKEATPEARGARRGVQKWVITDGESH 286

QY 119 SPOLKEVIOQSERNVTRYAVVLGYNNRGINGNPETPLNEIKVIASDPDDKHFFNVTD 178
DB :||:||||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 287 NYRLKQVIDCEDENIORFSIALULGHYNRGNLSLTERKEFVEEIKSIASEPTKHFVN 346

QY 179 ALKDIVDALGDRIEFSLEGT 197
DB |||:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 347 ALVTIVALKGERIFALEAT 365

RESULT 3
A55348
Integrin alpha-1 - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 15-Sep-2003
C;Accession: A55348
R;Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.
J. Biol. Chem. 269, 22811-22816, 1994
A;Title: The role of the I domain in ligand binding of the human integrin alpha-lbeta-1
A;Reference number: A55348; MUID:94357930; PMID:7521332
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-272 <KER>
A;Cross-references: GB:IU0114
F;55-230/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 58.5%; Score 598; DB 2; Length 272;
Best Local Similarity 57.3%; Pred. No. 3.3e-42;
Matches 114; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 CQTMDIVIVLDGNSNIYPWVEVQHFLINILKKFYIGPGQIQGVVGVEDVVHFEHLND 60
DB 52 CKTQLDIIIVLDGNSNIYPWESVTFANSLRNMDIGPQQTVGVIGVGTVVHFEYLNT 111

QY 61 YRSVKDVWEARASHIEQRGGTETRTAGIETPAREAFQK--GGRKGAKKMIVITDGESH 118
DB 112 YSSTEELVDAAIRLRGGGTCTMTALGIDTARBEEATEAHGARRGVQKWVITDGESH 171

QY 119 SPDLEKVIQSQERNVTRYAVVLGYNNRGINDPETPLNEIKYTASDPDDKHFFNVTD 178
DB 172 NYRLQEVLDKEDENIORFAILLGSYSGNLSTERKEFVEEIKSIASEPTKHFVN 231

QY 179 ALKDIVDALGDRIEFSLEGT 197
DB 232 ALVTIVALKGERIFALEAT 250

RESULT 4
S44142
VLA-2 protein homolog - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Sep-2003
C;Accession: S44142
R;Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latze
submitted to the EMBL Data Library, January 1994
Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not

```

A:Reference number: S44142
A:Accession: S44142
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1178 <DE>
A:Cross-references: EMBL:Z29987; NID:q473098; PIDN:CAA82877.1; PID:q473099
F:169-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match          47.6%; Score 486; DB 2; Length 1178;
Best Local Similarity 47.7%; Pred. No. 3.7e-32;
Matches 95; Conservative 37; Mismatches 65; Indels 2; Gaps 1

Qy      1 CQTMDIVIVLDGNSIYPWVEVCHFLINILKKFYIGPGQIQGVVQYGEDVWHEFLND 60
Db      166 CPSLDDVVVVCDENSIYPWEAVKQNFVLTGIDIGPKKTQVALIQYANERIIIFNLND 225

Qy      61 YRSVKDYVEAAASHTEQRGGTETRTAFAGIEFARSAFQK--GGRKGAKKMWIVITDGESH 118
Db      226 FETKEDVMVQATSETRQHGGLDITNFAIEFARDYAYSQTSGRPGATKVMVWVTDGESH 285

Qy      119 SPLEKVIQQSERDNNVRYAVAVGVYNNRGINPETFLNEIKYTASDPDKKHEFNVTDEA 178
Db      286 GSKLKVTIQQCNDEILRFGLAVGLYNNRNLDTNKLIIKEIKATASTETERYFFNVADEA 345

Qy      179 ALKDIDVALGDRIEFSLEGT 197
Db      346 ALLEKAGTLGQIFSIEST 364

RESULT 5
I45914
integrin alpha 2 subunit - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Sep-2003
C:Accession: I45914
R:Kamata, T.; Puzon, W.; Takada, Y.
J. Biol. Chem. 269, 9659-9663, 1994
A:Title: Identification of putative ligand binding sites within the I-domain
A:Reference number: A54402; MUID:94193647; PMID:7511592
A:Accession: I45914
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1170 <KAM>
A:Cross-references: GB:I25886; NID:g439695; PIDN:AAB59255.1; PID:g439696
F:161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

```

Query Match	45.8%	Score 468;	DB 2;	Length 1170;
Best Local Similarity	45.2%	Pred. No. 1.le-30;		
Matches	90;	Conservative 43;	Mismatches 64;	Indels 2; Gaps 1
QY	1	CQTYMDIVILVDGNSIYPVVEVOHFLINILKFFVIGPOIQVGVVQYGEDVVVHEFLND	60	
Db	158	CPSTIDVVVUCDENSIIYPDVAVKXFLKFFVQGLDIGTQKGLIQANNPRVVFNLT	217	
QY	61	YRSVKDVVEAAASHIEQRGGTTRTAFGIEFARSBAFQ--KGGKGAKKMVIIVITGESH	118	
Db	218	FKSKDEMIKATSKTQFYGGDITNTFKAIQYARDTAYSTAAGRPGATKMVWVVTGESH	277	
QY	119	SPDLEKVTQOSERDNVTFYAVVGYGNRRGINPETFLNEIKYIASDPDDKHFNVTDEA	178	
Db	278	GSUKAVIDQCNKNIILRFGFVAVLYLNRLNADTKNLKEIKAIASITERHFFNVSD	337	
QY	179	ALKDIVDALGDRIFSL	197	
Db	338	DLLEKAGTIGROI	356	

RESULT 6
A33998
integrin alpha-2 chain precursor - human
N:Alternate names: CD49b, platelet glycoprotein GPIa; VLA-2/collagen receptor
C:Species: Homo sapiens (man)
C:Date: 30-Mar-1990 #sequence revision 18-Sep-1992 #text change 15-Sep-2003

C;Accession: A33998; B56793; A53117
J;Takada, Y.; Hemler, M.E.
J. Cell Biol. 109, 397-407, 1989
A;Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit (platelet)
A;Reference number: A33998; MUID:69308879; PMID:2545729
A;Accession: A33998
A;Molecule type: mRNA
A;Residues: 1-1181 <TAK>
A;Cross-references: GB:X17033; NID:g33906; PIDN:CAA34894.1; PID:g33907
A;Note: the authors translated the codon GAT for residue 802 as Gln, GTC for residue 803
B;Catinel, B.; Parmentier, S.; Leung, L.L.; McGregor, J.L.
Biochem. J. 279, 419-425, 1991
A;Title: Separation of important new platelet glycoproteins (GP1a, GP1c, GPIIb and
A;Reference number: A56793; MUID:92061944; PMID:1953640
A;Accession: B56793
A;Molecule type: protein
A;Residues: 30-43 <CAT>
A;Experimental source: platelet
R;Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Tsung, Y.L.; Gafford, A.
J. Biol. Chem. 269, 463-469, 1994
A;Title: The human alpha-2 integrin gene promoter. Identification of positive and negative
A;Reference number: A53117; MUID:94103255; PMID:8276836
A;Accession: A53117
A;Molecule type: DNA
A;Residues: 1-16, 'V', 18-21 <ZUT>
A;Cross-references: GB:L24121; NID:g400342; PIDN:AAA16619.2; PID:g4583535
A;Note: authors translated the codon GTA for residue 17 as Leu
C;Genetics:
A;Gene: GB:ITGA2; CD49B
A;Cross-references: GDB:I28031; OMIM:192974
A;Map position: 5q11.1-5q11.2
C;Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-1133/Domain: extracellular #status predicted <EXT>
F;1134-1154/Domain: von Willebrand factor type A repeat homology <WMA2>
F;1155-1181/Domain: transmembrane #status predicted <TM>
F;1182-1200/Domain: intracellular #status predicted <CYT>
F;1201-1234/Domain: carbohydrate (Asn) (covalent)
Query Match 45.58; Score 465; DB 2; Length 1181;
Best Local Similarity 45.74; Pred. No. 2.1e-30;
Matches 91; Conservative 41; Mismatches 65; Indels 2; Gaps 1;
QY 1 CQTMDIVIVLDGNSIYPWVEQHFILNKKFYIGPQIQGVQYQGVGVVHEFLND 60
Db 169 CPSLIDVVVDCDSNIYPWDVKNFKLEFVQGLDGTQVGLIYANPRVFNLT 228
QY 61 YRSVKOVVEASHIEQRGGTETRTAFGEIFARSEAFQ--KGRKGAKKVMIVITDGHSD 118
Db 229 YKTEEMIVATSQTSQYGGDLTNTFGAIQYARKYVSAAGSRSAATKVMVVVDGESH 288
QY 119 SPLEKVIQOSERDNTVRYAVVLGYNRRGINPETFLNKKIYASDPDKHFNFVDEA 178
Db 289 GSMKKAVIDQNDHNIIRFGIYLVNRLDKNLKEIKAIASIPTRYFFNVSD 348
QY 179 ALKDIVDALGDRIFSLEGT 197
Db 349 ALLEKAGTLGEQIFSIEGT 367
RESULT 7
I47230
VLA-2 protein - pig (fragment)
N;Alternate names: glycoprotein Ia/Iia
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 15-Sep-2003
C;Accession: I47230; S21518
R;Bahou, W.F.; Potter, C.L.; Mirza, H.
Blood 84, 3734-3741, 1994
A;Title: The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific recognition
A;Reference number: I47230; MUID:95036279; PMID:7949129
A;Accession: I47230
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA
A;Residues: 1-191 <BAH>
A;Cross-references: EMBL:Z12137; NID:g2158; PIDN:CAA78125.1; PID:g2159
C;Keywords: Glycoprotein
Query Match 39.9%; Score 408; DB 2; Length 191;
Best Local Similarity 45.6%; Pred. No. 1.2e-26;
Matches 82; Conservative 35; Mismatches 61; Indels 2; Gaps 1;
QY 20 WVEVQHFILNKKFYIGPQIQGVVQYQGVGVVHEFLNDYRSVKDWEAASHIEQRGG 79
Db 1 WDAVKNFLEKFGVQGLDGTQVGLIYANPRVFNLTNFKTKAEWVEATSHTTQVGG 60
QY 80 TETRTAFGEIFARSEAFQ--KGRKGAKKVMIVITDGHSDSPLEKVIQOSERDNTVRY 137
Db 61 DLNTFKAIQYARDSAYSAAGRPGATKVMVVVDGESHGSMKKAVIDQNDHNIIR 120
QY 138 AVAVLGYNRRGINPETFLNKKIYASDPDKHFNFVDEAALKDIVDALGDRIFSLEGT 197
Db 121 GIAVLGYLVNRLDKNLKEIKAIASIPTRYFFNVSDADLLEKAGTLGEQIFSIEGT 180
RESULT 8
A40020
collagen alpha 1(XII) chain precursor - chicken
N;Alternate names: fibrochimerin
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Sep-2003
C;Accession: A40020; A34485; A28037; S23814; S22254; S28811
R;Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, Y.; Obai
J. Cell Biol. 115, 209-221, 1991
A;Title: The complete primary structure of type XII collagen shows a chimeric molecule
ous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site
A;Reference number: A40020; MUID:92011862; PMID:1918137
A;Accession: A40020
A;Molecule type: mRNA
A;Residues: 1-3124 <YAM>
A;Cross-references: GB:D00824; NID:g222810; PIDN:BAA00701.1; PID:g222811
R;Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
J. Biol. Chem. 264, 19772-19778, 1989
A;Title: Type XII collagen. A large multidomain molecule with partial homology to type
A;Reference number: A34485; MUID:90062079; PMID:2584192
A;Accession: A34485
A;Molecule type: mRNA
A;Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2876, 'F', 2978-3124 <GOR>
A;Cross-references: EMBL:J05137; NID:g211284; PIDN:AAA48635.1; PID:g211285
A;Accession: B34485
A;Molecule type: protein
A;Residues: 2772-2792; 2846-2873 <GOR2>
R;Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
A;Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA
A;Reference number: A28037; MUID:87317590; PMID:3476925
A;Accession: A28037
A;Molecule type: mRNA
A;Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
A;Cross-references: EMBL:ML7375; NID:g211649; PIDN:AAA48718.1; PID:g211650
A;Note: this sequence has been revised in reference A34485
R;Koch, M.; Bernasconi, C.; Chiquet, M.
Eur. J. Biochem. 207, 847-856, 1992
A;Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of
A;Reference number: S23814; MUID:92362621; PMID:1323460
A;Accession: S23814
A;Molecule type: protein
A;Residues: 'X', 1333, 'Q', 1335-1347; 1914-1928; 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517
R;Dublet, B.; van der Rest, M.
J. Biol. Chem. 262, 17724-17727, 1987
A;Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-
A;Reference number: S22254; MUID:88087065; PMID:3121603
A;Accession: S22254
A;Molecule type: protein
A;Residues: 2831-2832, 'T', 2834, 'R', 2836-2843; 3002-3014 <DUB>

F;148-318/Domain: von Willebrand factor type A repeat homology <VWA2>
F;1106-1129/Domain: transmembrane #status predicted <TM>

Query Match 23.5%; Score 240.5; DB 2; Length 1153;
Best Local Similarity 32.8%; Pred. No. 7.9e-12;
Matches 66; Conservative 41; Mismatches 81; Indels 13; Gaps 7;

QY 2 QTYMDIVILVDGNSI--YPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVHFEHLN 59
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 59
QY 146 QESDVLFLDGGSSINNDIPQKWEFVSIVMEQF--KKSKTLFSLMWOYSDEFFRIHFTFN 203
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 203
QY 60 DYRSKVDVWEAASHIEORGGTETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGESEH 117
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 117
QY 204 DFKNPSFRSHVSPKQLNG-RTKTASGIRKVVRELPHKTINGARENAKILVITDGEKF 262
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 262
QY 118 DSP-DLEKVIQQSERDNNVTRYAVAVLGYNRRGINPETFLNEIKYIASDDDDKHFFNVTD 176
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 176
QY 263 GDPDYKDVPIPEADRGVIRYVIGVGNFKNK---PQS-REELDTIASKPAGEVHFQVDN 317
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 317
QY 177 EALKDIDVDALGDRIFSLGFT 197
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 197
QY 318 FEALNTTQNLQEKIFALEGT 338
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 338

RESULT 11

I51027
type XII collagen alpha-1 chain - eastern newt (fragment)
A;Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 15-Sep-2003
C;Accession: I51027
Dev. Biol. 168, 503-513, 1995
A;Title: Monoclonal antibody M22 identifies the urodele alpha 1 chain of type XII collagen
A;Reference number: I51027; MUID:95246925; PMID:7729585
A;Accession: I51027
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-929 <WEI>
A;Cross-references: EMBL:U19494; NID:g632647; PIDN:AAA80217.1; PID:g632648
F;155-236/Domain: fibronectin type III repeat homology <3R>
F;631-795/Domain: von Willebrand factor type A repeat homology <VWA3>

Query Match 23.2%; Score 237.5; DB 2; Length 929;
Best Local Similarity 31.6%; Pred. No. 1.1e-11;
Matches 61; Conservative 49; Mismatches 60; Indels 23; Gaps 7;

QY 6 DIVILVDGNSI--YPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVHFEHLNDYRS 63
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 63
QY 633 DIVLLVDGWSIGRPNFKIVRNFTSRVVEVFDIGSDRVQIAVSQYSGDPRTEQLNTHKT 692
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 692
QY 64 VKDVEAASHIEORGGTETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGESEHSPD 121
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 121
QY 693 KKSIMDAVNLFPYKGG-NTWTGALKKEILENNFRPGVGMREKARKIAILLTDKQSD-- 749
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 749
QY 122 LEKVIQQSER--DNVTRYAVAVLGYNRRGINPETFLNEIKYIASDDDDKHFFNVNDEA 178
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 178
QY 750 ---IVAPSKRYADEGIELYAVGK-----NADE--NELKEIASDPDELYMNVADFS 796
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 796
QY 179 ALKDIDVDALGDRI 191
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 191
QY 797 LLTNVNDLTENV 809
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 809

RESULT 12

A45974
collagen alpha 1(XIV) chain precursor, short form 2 - chicken
N;Alternate names: undulin
C;Species: Gallus gallus (chicken)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Sep-2003
C;Accession: A45974; S30085; S22916; S17035; S20833
R;Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Llin
J. Biol. Chem. 268, 12177-12184, 1993
A;Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region

ns.

A;Reference number: A45974; MUID:93280195; PMID:8505337
A;Accession: A45974
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-1747 <GER>
A;Experimental source: embryo skin
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBIP:133365)
R;Apte, S.S.

submitted to the EMBL Data Library, March 1992

A;Reference number: S30085
A;Accession: S30085
A;Molecule type: mRNA
A;Residues: 1472-1660 <APT>
A;Cross-references: EMBL:X65122; NID:g62871; PIDN:CAA46238.1; PID:g938175
R;Trueb, J.; Trueb, B.
Eur. J. Biochem. 207, 549-557, 1992
A;Title: Type XIV collagen is a variant of undulin
A;Reference number: S22916; MUID:92339443; PMID:1339349
A;Accession: S22916
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>
R;Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemayer, T.F.; van der Rest, M.; Mayne, Eur. J. Biochem. 201, 333-338, 1991

A;Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens
A;Reference number: S17035; MUID:92037585; PMID:1935930
A;Accession: S17035
A;Molecule type: mRNA
A;Residues: 1472-1659 <GOR1>
A;Accession: S20833

A;Molecule type: protein

A;Residues: 1551-1570;1593-1599;1639-1667 <GOR2>

C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
F;40-204/Domain: von Willebrand factor type A repeat homology <VWA1>
F;236-317/Domain: fibronectin type III repeat homology <FN3A>
F;326-409/Domain: fibronectin type III repeat homology <FN3B>
F;418-498/Domain: fibronectin type III repeat homology <FN3C>
F;507-591/Domain: fibronectin type III repeat homology <FN3D>
F;625-707/Domain: fibronectin type III repeat homology <FN3E>
F;716-798/Domain: fibronectin type III repeat homology <FN3F>
F;806-893/Domain: fibronectin type III repeat homology <FN3G>
F;924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>
F;1111-1332/Domain: non-collagenous NC4 #status predicted <NC4>
F;1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>
F;1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 23.0%; Score 235.5; DB 2; Length 1747;
Best Local Similarity 32.8%; Pred. No. 3.4e-11;
Matches 63; Conservative 37; Mismatches 74; Indels 19; Gaps 6;

QY 1 CQT--YMDIVILVDGNSI--YPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVHFEH 56

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 56

QY 35 CKTPEATADIVILVDGWSIGRPNFKIVRNFTSRVVEVFDIGSDRVQIAVSQYSGDPRTEQLNTHKT 94

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 94

QY 57 HNDYRSKVDVWEAASHIEORGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESEH 114

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 114

QY 95 HLNAYGTKDAVLADRNLPYKGG-NTLTGLALTLYLENCFKPEAGARPGVSKIGILITDG 153

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 153

QY 115 ESHDSPLDKVIOQSERDNNVTRYAVAVLGYNRRGINPETFLNEIKYIASDDDDKHFFNV 174

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 174

QY 154 KQSD--DVTPPKNLEAGIELFAIGV-----KNADINELKEIASPEPDSTHYNV 201

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 201

QY 175 TDEAALKKDIDVDAL 187

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 187

QY 202 ADFNFNSIVEGL 214

Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 214

RESULT 13

S31212
collagen alpha 1(XIV) chain precursor, short form - chicken
C;Species: Gallus gallus (chicken)

C;Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 15-Sep-2003
C;Accession: S31212
R;Waelchli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A;Title: Complete primary structure of chicken collagen XIV.
A;Reference number: S31211; MUID:93185668; PMID:8444186
A;Accession: S31212
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-1857 <WAE>
A;Cross-references: EMBL:X70792; NID:G288874; PIDN:CAA50063.1; PID:G288875
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Genetics:
A;Gene: Col14A1

C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>
F;29-110/Domain: fibronectin type III repeat homology <FN3A>
F;156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F;352-433/Domain: fibronectin type III repeat homology <FN3B>
F;442-525/Domain: fibronectin type III repeat homology <FN3C>
F;534-614/Domain: fibronectin type III repeat homology <FN3D>
F;623-707/Domain: fibronectin type III repeat homology <FN3E>
F;741-823/Domain: fibronectin type III repeat homology <FN3F>
F;832-914/Domain: fibronectin type III repeat homology <FN3G>
F;922-1009/Domain: fibronectin type III repeat homology <FN3H>
F;1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>
Query Match 22.9%; Score 234.5; DB 2; Length 1857;
Best Local Similarity 33.7%; Pred. No. 4.5e-11;
Matches 66; Conservative 37; Mismatches 68; Indels 25; Gaps 8;
QY 1 CQT--YMDIVIVLDGNSI--YPMVEVQHFLINILKXFYIGPGQIQGVGVQGEDVVEHF 56
Db 151 CKTFAIADIIVLDGNSIGRFNLFVLFLFLENLVSFAFNVGSEKTRVGLAQYSGDPRIEW 210
QY 57 HLNDRYSKDVVEAASHIEQGGTETRTAFGIEFARSAF--QKGRKGAKKMWIVITDG 114
Db 211 HLNAYGTDKDAVLDAVRNLPYKGG--NTLTGLALTYLENSFPKEAGARPGVSKIGILITDG 269
QY 115 ESHDSPDLKVIQOSE--RD-NVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKH 171
Db 270 KSQDD-----VIPPAKNLRDAGIELFAIGV-----KNADINELKEIASBPDSTHV 314
QY 172 FNVTEAALKDIVDAL 187
Db 315 YNVADFNFMSIVEGL 330

RESULT 14

S78476
collagen alpha 1(XIV) chain precursor, long form - chicken
C;Species: Gallus gallus (Chicken)
C;Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 15-Sep-2003
C;Accession: S78476; S31211
R;Trueb, B.
submitted to the EMBL Data Library, January 1993
A;Reference number: S78476
A;Accession: S78476
A;Molecule type: mRNA
A;Residues: 1-1888 <TRU>
A;Cross-references: EMBL:X70793; NID:G288872; PIDN:CAA50064.1; PID:G288873
R;Waelchli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A;Title: Complete primary structure of chicken collagen XIV.
A;Reference number: S31211; MUID:93185668; PMID:8444186
A;Accession: S31211
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-416;1460-1811,1843-1898 <WAE>
A;Cross-references: EMBL:X70793
C;Genetics:
A;Gene: Col14A1

C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trim
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>
F;29-110/Domain: fibronectin type III repeat homology <FN3A>
F;156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F;352-433/Domain: fibronectin type III repeat homology <FN3B>
F;442-525/Domain: fibronectin type III repeat homology <FN3C>
F;534-614/Domain: fibronectin type III repeat homology <FN3D>
F;623-707/Domain: fibronectin type III repeat homology <FN3E>
F;741-823/Domain: fibronectin type III repeat homology <FN3F>
F;832-914/Domain: fibronectin type III repeat homology <FN3G>
F;922-1009/Domain: fibronectin type III repeat homology <FN3H>
F;1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 22.9%; Score 234.5; DB 2; Length 1898;
Best Local Similarity 33.7%; Pred. No. 4.5e-11;
Matches 66; Conservative 37; Mismatches 68; Indels 25; Gaps 8;
QY 1 CQT--YMDIVIVLDGNSI--YPMVEVQHFLINILKXFYIGPGQIQGVGVQGEDVVEHF 56
Db 151 CKTFAIADIIVLDGNSIGRFNLFVLFLFLENLVSFAFNVGSEKTRVGLAQYSGDPRIEW 210
QY 57 HLNDRYSKDVVEAASHIEQGGTETRTAFGIEFARSAF--QKGRKGAKKMWIVITDG 114
Db 211 HLNAYGTDKDAVLDAVRNLPYKGG--NTLTGLALTYLENSFPKEAGARPGVSKIGILITDG 269
QY 115 ESHDSPDLKVIQOSE--RD-NVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKH 171
Db 270 KSQDD-----VIPPAKNLRDAGIELFAIGV-----KNADINELKEIASBPDSTHV 314
QY 172 FNVTEAALKDIVDAL 187
Db 315 YNVADFNFMSIVEGL 330

RESULT 15

RWHUIC
cell surface glycoprotein CD11c precursor - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 22-Jun-1999
C;Accession: A36584; A35543; S00864
R;Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
J. Biol. Chem. 265, 12750-12751, 1990
A;Reference number: A36584
A;Contents: erratum
A;Accession: A36584
A;Molecule type: DNA
A;Residues: 1-1163 <COR>
A;Note: this revision to the sequence from reference A35543 includes the carboxyl end
R;Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
J. Biol. Chem. 265, 2782-2788, 1990
A;Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule
A;Reference number: A35543; MUID:90153906; PMID:2303426
A;Accession: A35543
A;Molecule type: DNA
A;Residues: 1-834 <CO2>
A;Note: this sequence has been revised in reference A36584
R;Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.
EMBO J. 6, 4023-4028, 1987
A;Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte
A;Reference number: S00864; MUID:88166645; PMID:3327687
A;Accession: S00864
A;Molecule type: mRNA
A;Residues: 1-755, '1', 757-1163 <CO3>
A;Cross-references: GB:M81695; EMBL:Y00093; NID:g487829; PIDN:AAA59180.1; PID:g487830
A;Note: part of this sequence was confirmed by protein sequencing
C;Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on m
C;Genetics:
A;Gene: GDB:ITGAX; CD11C
A;Cross-references: GDB:119758; OMIM:151510
A;Map position: 16p11.2-16p11.2
C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom

C:Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>
F:20-1167/Domain: extracellular #status predicted <EXT>
F:149-319/Domain: von Willebrand factor type A repeat homology <VWA4>
F:1108-1133/Domain: transmembrane #status predicted <TM>
F:1134-1163/Domain: intracellular #status predicted <INT>
F:61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match	21.9%	Score 223.5;	DB 1;	Length 1163;
Best Local Similarity	32.0%	Pred. No. 2e-10;		
Matches	63;	Conservative 36;	Mismatches 85;	Indels 13; Gaps 6;
QY	6	DIVIVLDSNSIYP--WVEVOHFLINILKFTYIGPGQIQGVGVQYGEDVWHEFHLDNYRS	63	
Db	151	DIVFLIDSGSGISSENATMNFVRAVISQFQ--RPSTQFSLMQFSNKFQTHFTFEFR	208	
QY	64	VKDVEEAASHIEQ-RGGTETATFGIEFARSEAFQKGGKGAKKVMIVITDGESH-DSPD	121	
Db	209	TSNPLSLIASVHQLQGFTYTATAIQNVVHRLFHASYGARRDATKILIVITDGKXGDSLD	268	
QY	122	LEKVIQOSERDNTRYAVAV-LGYNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL	180	
Db	269	YKDVIPADAAAGITRYALGVGLAFQNRNS-----WKELNDIASKPSQEHIFKVEDFDAL	322	
QY	181	KDIVDALGDRIFSLEGT	197	
Db	323	KDIQNQLKEKIFAIEGT	339	

Search completed: September 21, 2004, 13:05:10
Job time : 11.905 secs

This page Blank (usps10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:54:33 ; Search time 10.4553 Seconds
(without alignments)
972.742 Million cell updates/sec

Title: US-09-980-403-2_COPY_159_355

Perfect score: 1022

Sequence: 1 CQYMDIVILDGNSIYPW.....AALKDIVDALGDRIFSLEGT 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	594	58.1	214	4	US-10-061-658-6
2	594	58.1	214	4	US-10-061-658-9
3	593	58.0	214	4	US-10-061-658-5
4	507.5	49.7	1183	4	US-09-532-3108-5
5	454.5	44.5	1183	4	US-09-532-3108-6
6	244.5	23.9	1178	1	US-08-199-776-2
7	244.5	23.9	1178	3	US-08-663-731-2
8	244.5	23.9	1178	3	US-08-879-338-2
9	244.5	23.9	1178	5	PCT-US95-02044-2
10	244.5	23.9	1179	4	US-09-293-238B-2
11	238.5	23.3	413	1	US-08-485-618-101
12	238.5	23.3	413	2	US-08-605-672-101
13	238.5	23.3	413	2	US-08-482-293A-101
14	238.5	23.3	413	2	US-08-943-363-101
15	238.5	23.3	413	3	US-09-193-043-101
16	238.5	23.3	413	4	US-09-688-307A-101
17	238.5	23.3	413	4	US-09-350-259-101
18	223.5	21.9	1163	2	US-08-476-062A-44
19	223.5	21.9	1163	5	PCT-US96-01314-44
20	221.5	21.7	1163	1	US-08-173-497-4
21	221.5	21.7	1163	1	US-08-286-889-4
22	221.5	21.7	1163	1	US-08-485-618-4
23	221.5	21.7	1163	1	US-08-362-652-4
24	221.5	21.7	1163	2	US-08-605-672-4
25	221.5	21.7	1163	2	US-08-482-293A-4
26	221.5	21.7	1163	2	US-08-943-363-4
27	221.5	21.7	1163	3	US-09-193-043-4

```

28 221.5 21.7 1163 4 US-09-688-307A-4 Sequence 4, Appli
29 221.5 21.7 1163 4 US-09-350-259-4 Sequence 4, Appli
30 220.5 21.6 1151 1 US-08-286-889-37 Sequence 37, Appl
31 220.5 21.6 1151 1 US-08-485-618-37 Sequence 37, Appl
32 220.5 21.6 1151 1 US-08-362-652-37 Sequence 37, Appl
33 220.5 21.6 1151 2 US-08-605-672-37 Sequence 37, Appl
34 220.5 21.6 1151 2 US-08-482-293A-37 Sequence 37, Appl
35 220.5 21.6 1151 2 US-08-943-363-37 Sequence 37, Appl
36 220.5 21.6 1151 3 US-09-193-043-37 Sequence 37, Appl
37 220.5 21.6 1151 4 US-09-688-307A-37 Sequence 37, Appl
38 220.5 21.6 1151 4 US-09-350-259-37 Sequence 37, Appl
39 220.5 21.6 1161 1 US-08-485-618-55 Sequence 55, Appl
40 220.5 21.6 1161 1 US-08-362-652-55 Sequence 55, Appl
41 220.5 21.6 1161 2 US-08-605-672-55 Sequence 55, Appl
42 220.5 21.6 1161 2 US-08-482-293A-55 Sequence 55, Appl
43 220.5 21.6 1161 2 US-08-943-363-55 Sequence 55, Appl
44 220.5 21.6 1161 3 US-09-193-043-55 Sequence 55, Appl
45 220.5 21.6 1161 4 US-09-688-307A-55 Sequence 55, Appl

```

ALIGNMENTS

RESULT 1

```

US-10-061-658-6
; Sequence 6, Application US/10061658
; Patent No. 6652856
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koreliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
; CURRENT APPLICATION NUMBER: US/10/061,658
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
; US-10-061-658-6

```

```

Query Match 58.1%; Score 594; DB 4; Length 214;
Best Local Similarity 58.4%; Pred. No. 9.1e-61;
Matches 115; Conservative 33; Mismatches 47; Indels 2; Gaps 1;

```

```

QY 1 CQYMDIVILDGNSIYPWVEVQHFLINLKKFYIGPGQIQGVVQYGVGVVHEFLND 60
Db 17 CSTQLDIVILDGNSIYPWDSVTAFLNLLKRMIDIGPKQTQGVIVQYGVNTHFNLNK 76
QY 61 YSVKDVVEAAHIEQGGTETRTAPGIEFARSEAF--OKGGRKGAKKVMIVITDGHSD 118
Db 77 YGSTEELVAAKKIVQGRGQTMTALGTDRTARKEAFTEARGARRGVKKVMIVITDGHSD 136
QY 119 SPDLKVIQOQSERDNNVTRVAVVLGVYNNRRGNINPEFLNEIKYIASDPDDKHFENVYDEA 178
Db 137 NHRLLKKVIQODEBNLQRFSAIILGNSYNGNLSTFKFVEEIKSIASEPTEKHFENVYDEU 196
QY 179 ALKDIVDALGDRIFSLE 195
Db 197 ALVTIVKTLGERIFALE 213

```

RESULT 2

```

US-10-061-658-9
; Sequence 9, Application US/10061658
; Patent No. 6652856
; GENERAL INFORMATION:

```

```
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Kotliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
; CURRENT APPLICATION NUMBER: US/10/061,658
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-061-658-9

Query Match      58.1%; Score 594; DB 4; Length 214;
Best Local Similarity 58.4%; Pred. No. 9.1e-61;
Matches 115; Conservative 33; Mismatches 47; Indels 2; Gaps 1;

QY 1 CQTYMDIVILVLDGNSIYPWVEVQHFLINILKFFVIGPQIQVGVVQYGEDVWHEFHND 60
Db 17 CSTQLDIVILVLDGNSIYPWDSVAFINLILKRMIDIGPKQIQVGVVQYGENVTHFNLNK 76
QY 61 YRSVKDVVEAAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESH 118
Db 77 YSSTEELVAAKKIVQRGQTMTALTGTARKEAFTARGARRGVKKVMIVITDGESH 136
QY 119 SPLEKVIQOSERDNTVRYAVAVLGYNNRGINPETFLEIKYIASDDPKHFFNVTD 178
Db 137 NHRLKKVIGDCEDENIQRFSAIILGYNNRGNLSTKFEVEIKSIASEPTEKHFNVSD 196
QY 179 ALKQIVDALGDRIFSL 195
Db 197 ALVTIVKALGERIFALE 213

RESULT 3
US-09-532-310B-5
; Sequence 5, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
; Claffey, Kevin P
; Detmar, Michael
; TITLE OF INVENTION: Method for inhibiting tumor
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq.
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: Bis-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-532-310B-5

Query Match      49.7%; Score 507.5; DB 4; Length 1183;
Best Local Similarity 52.8%; Pred. No. 1.3e-49;
Matches 105; Conservative 32; Mismatches 59; Indels 3; Gaps 2;

QY 1 CQTYMDIVILVLDGNSIYPWVEVQHFLINILKFFVIGPQIQVGVVQYGEDVWHEFHND 60
Db 169 CSTXLDIVILVLDGNSIYPWDSVT--ALNDLLKRMIDIGPKXTXVGIVXYGENVTHFNLNK 227
QY 61 YRSVKDVVEAAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESH 118
Db 228 YSSTEELVAAKKIVYRGRTMTALTGTARKEAFTARGARRGVKKVMIVITDGESH 287
QY 119 SPLEKVIQOSERDNTVRYAVAVLGYNNRGINPETFLEIKYIASDDPKHFFNVTD 178
Db 288 NHRLKKVIGDCEDENIQRFSAIILGYNNRGNLSTKFEVEIKSIASEPTEKHFNVSD 347
QY 179 ALKQIVDALGDRIFSL 197
```

```

Db      348 ALVTIVKTLGERIFALEAT 366

RESULT 5
US-09-532-310B-6
; Sequence 6, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
;           Detmar, Michael
;           Claifey, Kevin P
; TITLE OF INVENTION: Method for inhibiting tumor
;                   angiogenesis in a living subject
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, Esq.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq.
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BIS-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-532-310B-6

Query Match      44.5%; Score 454.5; DB 4; Length 1183;
Best Local Similarity 45.5%; Pred. No. 1.9e-43;
Matches 91; Conservative 41; Mismatches 65; Indels 3; Gaps 2;

QY      1 CQYMDIVIVLDGNSIYPVVEVQHFLNLIKFFYIGPGQIQGVVQYGEDVVEHFLND 60
Db      169 CPSLIDVVVCDNSIYPWDVAVNFKLVQGLDGTQTKIQGLIQYANNPRVFNLT 228
QY      61 YRSVKDVEAASHIEQGGTETRTAFGIEFARSEAFQ--KGGRKGAKKVMIVITDGESH 118
Db      229 YKTEEMIVATSTQSYGGDLTNTFGAIQYARKYVSAASGRRSATKVMVVITDGESH 288
QY      119 SPLEKVIQQSERDNVTRYAVAVLGYNNRRGINPE--TFLNEIKVIASDPDDKHFFNV 177
Db      289 GSMLKAVIDQCNHILRFGVILGYLNRNALDTKMLIKEIKALASIPTERYFFNVSD 348
QY      178 AALKDIVDALGDRIFSLEGT 197
Db      349 AALLEKAGTLGEQIFSLEGT 368

RESULT 6
US-08-199-776-2
; Sequence 2, Application US/08199776
; Patent No. 5594120
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.

; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: No. 5594120el integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
US-08-199-776-2

Query Match      23.9%; Score 244.5; DB 1; Length 1178;
Best Local Similarity 33.7%; Pred. No. 4.4e-19;
Matches 68; Conservative 31; Mismatches 82; Indels 21; Gaps 6;

QY      6 DIVILVDGNSIYP--WVEVQHFLNLIKFFYIGPGQIQGVVQYGEDVVEHFLNDYRS 63
Db      201 EIAIILDGSGSIDPPDFQRAKDFISNMNRNFYKCFECNPFALVQYGGVITQTFDLRDSQ 260
QY      64 VKDVEAASHIEQGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHDP- 120
Db      261 VMASLARVQNVITQGVSV-TKTASAMQHVLDSITSSHGSRKASKVMVITDGIFEDPL 319
QY      121 DLEKVIQQSERDNVTRYAVAVLGYNNRRGINPETF-----LNEIKVIASDPDDKHFFNV 175
Db      320 NLATTVINSPKMQGVERFAIGV-----GEEFKSARTARELNLIASDPDETHAPKVT 369
QY      176 DEALKDIVDALGDRIFSLEGT 197
Db      370 NYMALDGLLSKRYNIISMEGT 391

RESULT 7
US-08-663-731-2
; Sequence 2, Application US/08663731
; Patent No. 6057423
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: No. 6057423el integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663,731
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,776
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1178 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-663-731-2

Query Match      23.9%; Score 244.5; DB 3; Length 1178;
Best Local Similarity 33.7%; Pred. No. 4.4e-19;
Matches 68; Conservative 31; Mismatches 82; Indels 21; Gaps 6;

QY 6 DIVVLGDSNIYP--WVEVQHFLINILKPYIGPGQIQGVGVQYGEDVWHEFHNDYRS 63
Db 201 EIAIILDGSGSIDPPDFQRAKDFISNMNRYEKCFCNCFALVQYGGVQIQTEFDLRSDQ 260

QY 64 VKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHDS- 120
Db 261 VNASLARVQNTQVGSV-TKTASAMQHVLDSIFTSSHGSRKASKVMVVLTDGGIFEDPL 319

QY 121 DLEKVIQOSERDNTRYAVAVLGYNRRGINPETP-----LNEIKYIASDPDDKHFFNVT 175
Db 320 NLTTVINSPKMQGVVERFAIGV-----GEEFKSARTARELNLIASDPDETHAFKVT 369

QY 176 DEALKKDIVDALGDRIFSLEGT 197
Db 370 NYMALDGLLSKLYNIISMEGT 391

RESULT 8
US-08-879-338-2
Sequence 2, Application US/08879338A
Patent No. 6063906
GENERAL INFORMATION:
APPLICANT: Brenner, Michael B.
APPLICANT: Parker, Christina M.
TITLE OF INVENTION: Antibodies to No. 6063906el Integrin Alpha
FILE REFERENCE: B0801/7080/ERP
CURRENT APPLICATION NUMBER: US/08/879,338A
CURRENT FILING DATE: 1997-06-20
EARLIER APPLICATION NUMBER: US 08/663,731
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: US 08/199,776
EARLIER FILING DATE: 1994-02-18
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1178
TYPE: PRT
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: (-18)...(-1)
US-08-879-338-2

Query Match      23.9%; Score 244.5; DB 3; Length 1178;
Best Local Similarity 33.7%; Pred. No. 4.4e-19;
Matches 68; Conservative 31; Mismatches 82; Indels 21; Gaps 6;

QY 6 DIVVLGDSNIYP--WVEVQHFLINILKPYIGPGQIQGVGVQYGEDVWHEFHNDYRS 63
Db 201 EIAIILDGSGSIDPPDFQRAKDFISNMNRYEKCFCNCFALVQYGGVQIQTEFDLRSDQ 260

QY 64 VKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHDS- 120
Db 261 VNASLARVQNTQVGSV-TKTASAMQHVLDSIFTSSHGSRKASKVMVVLTDGGIFEDPL 319

QY 121 DLEKVIQOSERDNTRYAVAVLGYNRRGINPETP-----LNEIKYIASDPDDKHFFNVT 175
Db 320 NLTTVINSPKMQGVVERFAIGV-----GEEFKSARTARELNLIASDPDETHAFKVT 369

QY 176 DEALKKDIVDALGDRIFSLEGT 197
Db 370 NYMALDGLLSKLYNIISMEGT 391

RESULT 9
PCT-US95-02044-2
Sequence 2, Application PC/TUS9502044
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Novel integrin alpha subunit
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Wolf, Greenfield and Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent'n Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02044
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,776
FILING DATE: 18 February 1994
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1178 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02044-2

Query Match      23.9%; Score 244.5; DB 5; Length 1178;
Best Local Similarity 33.7%; Pred. No. 4.4e-19;
Matches 68; Conservative 31; Mismatches 82; Indels 21; Gaps 6;

QY 6 DIVVLGDSNIYP--WVEVQHFLINILKPYIGPGQIQGVGVQYGEDVWHEFHNDYRS 63
Db 201 EIAIILDGSGSIDPPDFQRAKDFISNMNRYEKCFCNCFALVQYGGVQIQTEFDLRSDQ 260

QY 64 VKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHDS- 120
Db 261 VNASLARVQNTQVGSV-TKTASAMQHVLDSIFTSSHGSRKASKVMVVLTDGGIFEDPL 319
```

```
Qy 121 DLEKVIQSSRDNVTRVAVAVLGYNNRGINPETF-----LNEIKYIASDPDDKHFENVT 175
Db 320 NLTTVINSRQMGVERFAIGV-----GEEFKSARTARELNLIASDPDETHAFKVT 369
Qy 176 DEAAKDIVDALGDRIFSLRG 197
Db 370 NYWALDGLLSKLYNIISMEGT 391

RESULT 10
US-09-293-238B-2
; Sequence 2, Application US/09293238B
; Patent No. 6455042
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: A Method of Treating Ulcerative Colitis
; TITLE OF INVENTION: or Crohn's Disease by Administering an Antibody to Alpha E
; FILE REFERENCE: L0560/7005/ERP
; CURRENT APPLICATION NUMBER: US/09/293,238B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: US 08/879,338
; PRIOR FILING DATE: 1997-06-20
; PRIOR APPLICATION NUMBER: US 08/663,731
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: US 08/199,776
; PRIOR FILING DATE: 1994-02-18
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1179
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(18)
US-09-293-238B-2
```

```
Query Match 23.9%; Score 244.5; DB 4; Length 1179;
Best Local Similarity 33.7%; Pred. No. 4.4e-19;
Matches 68; Conservative 31; Mismatches 82; Indels 21; Gaps 6;

Qy 6 DIVIVLDGNSIYP--WVEVQHFLNLIKFKYIGPGQIQGVQVQYGEDVWHEFLNDYRS 63
Db 202 EIALILDGSGSIDPPDQRAKDFISNMNRNFEYKCFECFNALVQYGGVIOQTEFDLRDSQD 261
Qy 64 VKDVVFAASHIEQRGGTETRTAFGIEFARSEAP--QKGRKGAKKVMIVITDGESHDSP- 120
Db 262 VMASLARVQNIQVGSV-TKTASAMQHVLDSIFTSSHSGSRKASKVMVLTDDGIFEDPL 320
Qy 121 DLEKVIQSSRDNVTRVAVAVLGYNNRGINPETF-----LNEIKYIASDPDDKHFENVT 175
Db 321 NLTTVINSRQMGVERFAIGV-----GEEFKSARTARELNLIASDPDETHAFKVT 370
Qy 176 DEAAKDIVDALGDRIFSLRG 197
Db 371 NYWALDGLLSKLYNIISMEGT 392
```

```
RESULT 11
US-08-485-618-101
; Sequence 101, Application US/08485618
; Patent No. 5729533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
```

```
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,618
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32797
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 101:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-618-101
```

```
Query Match 23.3%; Score 238.5; DB 1; Length 413;
Best Local Similarity 34.8%; Pred. No. 4.5e-19;
Matches 69; Conservative 35; Mismatches 81; Indels 13; Gaps 6;

Qy 5 MDIVIVLDGNSI--YPWVEVQHFLNLIKFKYIGPGQIQGVQVQYGEDVWHEFLNDYR 62
Db 146 MDIVIVLDGSGSISSNDFRKXKDFRAVMDQF--KDTNTQFSLMQSNVLVTHFTSSPR 203
Qy 63 SVKDVVFAASHIEQRGGTETRTAFGIEFARSEAFQ--KGRKGAKKVMIVITDGESHDSP 120
Db 204 NSSNPQLVEPIVQLTGL-TFTATGILKVVVTELFQTKNGARESAKKILIVITDGQYKDP 262
Qy 121 -DLEKVIQSSRDNVTRVAVAVLGYNNRGINPETFELNEIKYIASDPDDKHFENVTDEAA 179
Db 263 LHYSVAPIQAEQAGIIRYAIQVGDFAQK-----PTARQELDTIASBPDAHFQVQDNFSA 317
Qy 180 LKDIVDALGDRIFSLRG 197
Db 318 LSSIQQLYDRIFAVEGT 335
```

```
RESULT 12
US-08-605-672-101
; Sequence 101, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
```

; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 101:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-605-672-101

Query Match 23.3%; Score 238.5; DB 2; Length 413;
Best Local Similarity 34.8%; Pred. No. 4.5e-19;
Matches 69; Conservative 35; Mismatches 81; Indels 13; Gaps 6;
QY 5 MDIVFLDGSNSI--YPWVEVQHFLINILKFFYIGPGQIQGVVQYGEDVVHEHFLNDYR 62
Db 146 MDIVFLDGSNSISSNDFRKMDFVRAVMDQF--KDTNTQFSLMQYSNVLVTHFTFSSFR 203
QY 63 SVKDVVEAASHIEQCGTETRTAFGIEFARSEAFQ--KGRKGAKVMIVITDGESHDP 120
Db 204 NSSNPQGLVEPIVQLTGL-TFTATGILKVVTELFTQKNGARES AKKILIVITDGQYKDP 262
QY 121 -DLEKVIQOSRDNRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAA 179
Db 263 LHYSAVIQAQAGIIRYAIGVGDAFOK-----PTAQELDTIASPPDAHVQVDNFA 317
QY 180 LKDIVDALGDRIFSLEGT 197
Db 318 LSSIQKQLYDRIFAVEGT 335

RESULT 13
US-08-482-293A-101
; Sequence 101, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,293A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 101:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-293A-101

Query Match 23.3%; Score 238.5; DB 2; Length 413;
Best Local Similarity 34.8%; Pred. No. 4.5e-19;
Matches 69; Conservative 35; Mismatches 81; Indels 13; Gaps 6;
QY 5 MDIVFLDGSNSI--YPWVEVQHFLINILKFFYIGPGQIQGVVQYGEDVVHEHFLNDYR 62
Db 146 MDIVFLDGSNSISSNDFRKMDFVRAVMDQF--KDTNTQFSLMQYSNVLVTHFTFSSFR 203
QY 63 SVKDVVEAASHIEQCGTETRTAFGIEFARSEAFQ--KGRKGAKVMIVITDGESHDP 120
Db 204 NSSNPQGLVEPIVQLTGL-TFTATGILKVVTELFTQKNGARES AKKILIVITDGQYKDP 262
QY 121 -DLEKVIQOSRDNRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAA 179
Db 263 LHYSAVIQAQAGIIRYAIGVGDAFOK-----PTAQELDTIASPPDAHVQVDNFA 317
QY 180 LKDIVDALGDRIFSLEGT 197
Db 318 LSSIQKQLYDRIFAVEGT 335

RESULT 14
US-08-943-363-101
; Sequence 101, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 12:44:02 ; Search time 5.77793 Seconds
(without alignments)
1775.344 Million cell updates/sec

Title: US-09-980-403-2_COPY_159_355

Perfect score: 1022

Sequence: 1 CQYMDIVILVDSGSIYPW.....AALKDIVDALGRIFSLG 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1022	100.0	1189	1	ITAH_HUMAN
2	599	58.6	1151	1	ITAH_HUMAN
3	599	58.6	1180	1	ITAH_RAT
4	598	58.5	285	1	ITAH_CHICK
5	590	57.7	1167	1	ITAG_HUMAN
6	486	47.6	1178	1	ITAG_MOUSE
7	468	45.8	1170	1	ITAG_BOVIN
8	465	45.5	1181	1	ITAH_HUMAN
9	262.5	25.7	3119	1	CA1C_MOUSE
10	255.5	25.0	3063	1	CA1C_HUMAN
11	248.5	24.3	3124	1	CA1C_CHICK
12	245.5	24.0	1167	1	ITAE_MOUSE
13	244.5	23.9	1179	1	ITAE_HUMAN
14	240.5	23.5	1153	1	ITAM_MOUSE
15	237.5	23.2	929	1	CA1C_NOTVI
16	234.5	22.9	1888	1	CA1C_CHICK
17	223.5	21.9	1163	1	ITAX_HUMAN
18	217	21.2	956	1	MTN2_MOUSE
19	216.5	21.2	1152	1	ITAM_HUMAN
20	211	20.6	956	1	MTN2_HUMAN
21	202.5	19.8	1162	1	ITAD_HUMAN
22	199.5	19.5	3137	1	CA36_CHICK
23	194.5	19.0	496	1	CAMA_HUMAN
24	193	18.9	500	1	CAMA_MOUSE
25	193	18.9	3176	1	CA36_HUMAN
26	189.5	18.5	493	1	CAMA_CHICK
27	188	18.4	2944	1	CA17_HUMAN
28	182	17.8	486	1	MTN3_HUMAN
29	176	17.2	3767	1	MUA3_CAEEL
30	174.5	17.1	619	1	MTN4_HUMAN
31	174	17.0	481	1	MTN3_MOUSE
32	173.5	17.0	1329	1	RF10_HUMAN
33	173	16.9	1170	1	ITAH_HUMAN

34	166	16.2	624	1	MTN4_MOUSE
35	163.5	16.0	639	1	CA1C_RABIT
36	162	15.9	452	1	MTN3_CHICK
37	152.5	14.9	547	1	COCH_CHICK
38	150.5	14.7	550	1	COCH_HUMAN
39	148.5	14.5	1163	1	ITAH_MOUSE
40	145.5	14.2	552	1	COCH_MOUSE
41	133.5	13.1	2813	1	VWF_CANFA
42	132.5	13.0	2813	1	VWF_HUMAN
43	126.5	12.4	1029	1	CA26_MOUSE
44	119	11.6	2482	1	VWF_PIG
45	117.5	11.5	1019	1	CA26_HUMAN

ALIGNMENTS

RESULT 1
ITAH_HUMAN
ID ITAH_HUMAN STANDARD; PRT; 1189 AA.
AC Q9UKX5; Q9UKQ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-11 precursor.
GN ITGAl1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Fetal heart, and Osteoblast;
RX MEDLINE=99417678; PubMed=10486209;
RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
Wang S.-X., Morris C.M., Krissansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel
human integrin alpha11 subunit (ITGAl1).";
RL Genomics 60:179-187(1999).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Fetal muscle, and Uterus;
RX MEDLINE=99395147; PubMed=10464311;
RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;
RT "CDNA Cloning and Chromosomal Localization of Human alpha(11)
Integrin. A collagen-binding, i domain-containing, beta(1)-associated
integrin alpha-chain present in muscle tissues.";
RL J. Biol. Chem. 274:25735-25742(1999).
[3]
RN SEQUENCE OF 954-1188 FROM N.A.
RP TISSUE=Fibroblast;
RA Andreu N., Estivill X., Escarceller M., Sumoy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
CC ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
CC HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,
CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
CC -!- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
CC FETAL MUSCLE CELLS (IN VITRO).
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -


```

DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00191; int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWF; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Repeat; Calcium; Magnesium; 3D-structure.
FT DOMAIN 1 1113 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1114 1136 POTENTIAL.
FT DOMAIN 1137 1151 CYTOPLASMIC (POTENTIAL).
FT REPEAT 16 75 FG-GAP 1.
FT REPEAT 2 2 FG-GAP 2.
FT DOMAIN 147 360 VWF_A.
FT REPEAT 349 404 FG-GAP 3.
FT REPEAT 405 457 FG-GAP 4.
FT REPEAT 459 520 FG-GAP 5.
FT REPEAT 520 599 FG-GAP 6.
FT REPEAT 602 654 FG-GAP 7.
FT CA_BIND 470 478 POTENTIAL.
FT CA_BIND 552 560 POTENTIAL.
FT CA_BIND 614 622 POTENTIAL.
FT SITE 1139 1142 GFFKR MOTIF.
FT DISULFID 54 64 BY SIMILARITY.
FT DISULFID 660 669 BY SIMILARITY.
FT DISULFID 675 728 BY SIMILARITY.
FT DISULFID 780 786 BY SIMILARITY.
FT DISULFID 850 858 BY SIMILARITY.
FT DISULFID 1002 1034 BY SIMILARITY.
FT DISULFID 1037 1044 BY SIMILARITY.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 671 671 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 720 720 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 812 812 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 855 855 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 880 880 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 887 887 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 946 946 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 980 980 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1085 1085 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 1151 AA; 127837 MW; 63F3C1AABF52808 CRC64;

Query Match 58.6%; Score 599; DB 1; Length 1151;
Best Local Similarity 58.3%; Pred. No. 6.3e-42;
Matches 116; Conservative 33; Mismatches 48; Indels 2; Gaps 1;

Qy 1 CQTMDIVLDGNSISYPWVEVQHFLINILKPYIGPQIQGVQYGVGDVHFEHLND 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 CSTQDIVLDGNSISYPWDSVTAFLNDLLKMDIGPKQTVQVIGYGVNVTHEFNLNK 198
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 61 YRSVKDVVEASHLEORGSTETRTAFGLFEARSAF--QKGRKGAKKMVLVIIDGESHD 118
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Dy 199 YSSTEELVLAARKIVQGRGQTMTALGDTATKEARTEARGARGVKKVMVITDGSND 258
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 119 SPDLKVIQSERDQNVTRYAVAVLGYNNRRGINPETFELNFKYIASDPDDKHFENVTDEA 178
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 259 NHRLKKVIOCEDENIQRFSAIILGNSYRGNLSTKTEKVEIKSIASEPTKHFNFVSD 318
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 179 ALKDIDVALGDRIPSLEGT 197
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 319 ALVTIVKTLGERIFALEAT 337
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
ITAL RAT STANDARD; PRT; 1180 AA.
AC P18614;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1) (CD49a).
DE ITGAL.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OK NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=90338125; PubMed=2380249;
RA Ignatius M.J., Large T.H., Houde M., Tawil J.W., Barton A.,
RA Esch F., Carbonetto S., Reichardt L.F.;
RT "Molecular cloning of the rat integrin alpha 1-subunit: a receptor
RT for laminin and collagen.";
RL J. Cell Biol. 111:709-720(1990).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.
RX MEDLINE=99313197; PubMed=10386626;
RA Nolte M., Pepinsky R.B., Vanyaminov S.Y., Kotliansky V.,
RA Gotwals P.J., Karpusas M.;
RT "Crystal structure of the alpha1beta1 integrin I-domain: insights into
RT integrin I-domain function.";
RL FEBS Lett. 452:379-385(1999).
CC -!- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-P-P-G-
CC E-R IN COLLAGEN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52140; CAA36384.1; -.
CC PIR; A35854; A35854.
CC PDB; 1CK4; 03-MAY-00.
CC InterPro; IPR000413; Integrin_alpha.
CC InterPro; IPR002035; VWF_A.
CC Pfam; PF01839; FG-GAP; 3.
CC Pfam; PF00357; integrin_A; 1.
CC Pfam; PF00092; vwa; 1.
CC SMART; SM00191; Int_alpha; 5.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC PROSITE; PS0234; VWF; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium; 3D-structure.
FT SIGNAL 1 28

```

```

FT CHAIN 29 1180 INTEGRIN ALPHA-1.
FT DOMAIN 29 1142 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1143 POTENTIAL.
FT DOMAIN 1146 1180 CYTOPLASMIC (POTENTIAL).
FT REPEAT 44 103 FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT DOMAIN 175 388 VWFA.
FT REPEAT 377 432 FG-GAP 3.
FT REPEAT 433 484 FG-GAP 4.
FT REPEAT 485 565 FG-GAP 5.
FT REPEAT 567 626 FG-GAP 6.
FT REPEAT 629 681 FG-GAP 7.
FT CA_BIND 497 505 POTENTIAL.
FT CA_BIND 579 587 POTENTIAL.
FT CA_BIND 641 649 POTENTIAL.
FT SITE 1168 1172 GFFKR MOTIF.
FT DISULFID 82 92 BY SIMILARITY.
FT DISULFID 687 696 BY SIMILARITY.
FT DISULFID 702 755 BY SIMILARITY.
FT DISULFID 807 813 BY SIMILARITY.
FT DISULFID 877 885 BY SIMILARITY.
FT DISULFID 1029 1062 BY SIMILARITY.
FT DISULFID 1066 1073 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 459 459 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1007 1007 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362EE4 CRC64;

Query Match 58.6%; Score 599; DB 1; Length 1180;
Best Local Similarity 57.8%; Pred. No. 6.4e-42;
Matches 115; Conservative 38; Mismatches 44; Indels 2; Gaps 1;

QY 1 CQTYMDIVILVDGNSIYPVVEVQHFLNLIKFKYIGPGQIQGVGVQYGEDVVFHFLND 60
Db 167 CSTQLDILVLDGNSIYPWESVIAFLNLLKRMIDIGPKQIQGVGVQYGENVTFEFLNK 226

QY 61 YRSVKDVVEAASHIEQCGGTETRTAFGIEFARSEAF--QKGRKAKKVMIVITDGESH 118
Db 227 YSSTEELVIAANKIGRQGLQMTALGIDTAREAEFTAEARGARRGVQKVMIVITDGESH 286

QY 119 SPDLKVTQQSERDENVTRYAVAVLGYNRRGINPETFLEIKYIASDPDDKHFFNVNDEA 178
Db 287 NYRLKQVLDQDCDENIQRSFIALGHYNGRLNSTEKFVEEIKSIASEFTEKHFFNVNDEL 346

QY 179 ALKDIVDALGDRIFSLGFT 197
Db 347 ALVTIVTIGALGERIFALEAT 365

RESULT 4
ITAL_CHICK STANDARD; PRT; 285 AA.
ID ITAL_CHICK

```

```

AC Q90615;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a)
DE (Fragment).
GN ITGAL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gizzard;
RX MEDLINE=94357930; PubMed=7521332;
RA Kern A., Briesewitz R., Bank I., Marcantonio E.E.;
RT "The role of the I domain in ligand binding of the human integrin
alpha 1 beta 1.";
RL J. Biol. Chem. 269:22811-22816(1994).
CC -!- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC E-R IN COLLAGEN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U01114; AAA59067.1; -.
DR HSP; P17301; IAOX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN ALPHA; PARTIAL.
DR PROSITE; PS0234; VWFA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane.
FT NON_TER 1 1
FT DOMAIN <1 >285 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 66 279 VWFA.
FT CARBOHYD 2 2 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 285 285
SQ SEQUENCE 285 AA; 31503 MW; 1B05D3246CD5CA7E CRC64;

Query Match 58.5%; Score 598; DB 1; Length 285;
Best Local Similarity 57.3%; Pred. No. 1.5e-42;
Matches 114; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

QY 1 CQTYMDIVILVDGNSIYPVVEVQHFLNLIKFKYIGPGQIQGVGVQYGEDVVFHFLND 60
Db 58 CKTQDILVLDGNSIYPWESVIAFLNLLKRMIDIGPKQIQGVGVQYGENVTFEFLNK 117

QY 61 YRSVKDVVEAASHIEQCGGTETRTAFGIEFARSEAFQK--QKGRKAKKVMIVITDGESH 118
Db 118 YSSTEELVIAANKIGRQGLQMTALGIDTAREAEFTAEARGARRGVQKVMIVITDGESH 177

QY 119 SPDLKVTQQSERDENVTRYAVAVLGYNRRGINPETFLEIKYIASDPDDKHFFNVNDEA 178
Db 178 NYRLQEVLDKDCDENIQRSFIALGHYNGRLNSTEKFVEEIKSIASEFTEKHFFNVNDEL 237

```

```

QY 179 ALKDIVDGLDRIFSLGCT 197
Db 238 ALVTIVLALGERIFALEAT 256

RESULT 5
ITAG HUMAN
ID ITAG_HUMAN STANDARD; PRT; 1167 AA.
AC O55578; Q9UH28;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-10 precursor.
GN ITGA10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular chondrocytes;
RX MEDLINE=98352078; PubMed=9685391;
RA Camper L., Hellman U., Lundgren-Aakerlund E.;
RT "Isolation, cloning, and sequence analysis of the integrin subunit
alpha10, a beta1-associated collagen binding integrin expressed on
chondrocytes.";
RL J. Biol. Chem. 273:20383-20389(1998).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells, and Heart;
RX MEDLINE=20169197; PubMed=10702680;
RA Lehnert K., Ni J., Ieung E., Gough S.M., Morris C.M., Liu D.,
Wang S.-X., Langley R., Krissansen G.W.;
RT "The integrin alpha10 subunit: expression pattern, partial gene
structure, and chromosomal localization.";
RL Cytogenet. Cell Genet. 87:238-244(1999).
CC -1- FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10
CC ASSOCIATES WITH BETA-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Widely expressed with highest expression in
muscle and heart. Found in articular cartilage.
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 1 VWFA domain.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF074015; AAC31952.1; -.
DR EMBL; AF112345; AAF21944.1; -.
DR EMBL; AF172723; AAF61638.1; -.
DR HSP; P17301; IAOX.
DR Genew; HGNC:6135; ITGA10.
DR MIM; 604042; -.
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0004895; F:cell adhesion receptor activity; TAS.
DR GO; GO:0005518; F:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.

```

SMART; SM00191; Int_alpha; 4.
SMART; SM00327; VWFA_1.
PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
PROSITE; PS0234; VWFA; 1.
KW Signal; Repeat; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1167 INTEGRIN ALPHA-10.
FT DOMAIN 23 1122 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1123 1145 POTENTIAL.
FT DOMAIN 1146 1167 CYTOPLASMIC (POTENTIAL).
FT REPEAT 38 97 FG-GAP 1.
FT REPEAT 2 2 FG-GAP 2.
FT DOMAIN 167 350 VWFA.
FT REPEAT 365 427 FG-GAP 3.
FT REPEAT 428 482 FG-GAP 4.
FT REPEAT 483 545 FG-GAP 5.
FT REPEAT 546 605 FG-GAP 6.
FT REPEAT 608 660 FG-GAP 7.
FT DOMAIN 1134 1140 POLY-LEU.
FT CA_BIND 494 502 POTENTIAL.
FT CA_BIND 558 566 POTENTIAL.
FT CA_BIND 620 628 POTENTIAL.
FT DISULFID 76 86 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 681 736 BY SIMILARITY.
FT DISULFID 789 795 BY SIMILARITY.
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 733 733 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 763 763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1011 1011 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1018 1018 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1039 1039 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 844 844 I -> L (IN REF. 2).
FT CONFLICT 909 909 G -> V (IN REF. 2).
FT CONFLICT 926 926 E -> D (IN REF. 2).
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AE0 CRC64;

Query Match 57.7%; Score 590; DB 1; Length 1167;
Best Local Similarity 58.3%; Pred. No. 3.6e-41;
Matches 116; Conservative 32; Mismatches 49; Indels 2; Gaps 1;

QY 1 CQYMDIVILDGNSGIYPMWVEVQHFLINILKKFYIGPGQIQGVVQYGVVHFEHLND 60
Db 162 CPTYMDVIVLDGNSGIYPMWSEVQTFRLRLVGLKFLDPEQIQVGLVQYGSFVHWSLGD 221
QY 61 YRSVKDVEAASHIEORGGTETATFIEFARSEAFQK--GGRGAKKVMIVITDGSMD 118
Db 222 FRTKEEVRAAKNLRSREGRETAKQAIMVACTEGFSQSHGGRPEARLLVWVTDGSHD 281
QY 119 SPLEKVIQOSERDNTYRAVAVLGYNNRGINPETFLEIKYIASDPPDKHFNVTDEA 178
Db 282 GEELPAALKACEAGRYTRYGIAVLHYLRQRDPSSFLREIRTIASDPDRFFFNVTDEA 341
QY 179 ALKDIVDGLDRIFSLGCT 197
Db 342 ALTDIVDGLDRIFSLGCT 360

RESULT 6
ITAG_MOUSE
ID ITAG_MOUSE STANDARD; PRT; 1178 AA.
AC Q62459; Q62163;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).


```
PT DOMAIN 813 903 FIBRONECTIN TYPE-III 5.
FT DOMAIN 904 998 FIBRONECTIN TYPE-III 6.
FT DOMAIN 999 1085 FIBRONECTIN TYPE-III 7.
FT DOMAIN 1086 1178 FIBRONECTIN TYPE-III 8.
FT DOMAIN 1199 1371 WFVA 3.
FT DOMAIN 1384 1473 FIBRONECTIN TYPE-III 9.
FT DOMAIN 1474 1564 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1565 1652 FIBRONECTIN TYPE-III 11.
FT DOMAIN 1654 1751 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1752 1842 FIBRONECTIN TYPE-III 13.
FT DOMAIN 1843 1932 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1933 2023 FIBRONECTIN TYPE-III 15.
FT DOMAIN 2024 2114 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2115 2202 FIBRONECTIN TYPE-III 17.
FT DOMAIN 2203 2291 FIBRONECTIN TYPE-III 18.
FT DOMAIN 2323 2496 WFVA 4.
FT DOMAIN 2520 2712 TSP N-TERMINAL.
FT DOMAIN 2451 2746 NONHELICAL REGION (NC3).
FT DOMAIN 2747 TRIPLE-HELICAL REGION
(COL2) WITH 1 IMPERFECTION.
FT DOMAIN 2899 2941 NONHELICAL REGION (NC2).
FT DOMAIN 2942 3044 TRIPLE-HELICAL REGION
(COL1) WITH 2 IMPERFECTIONS.
FT DOMAIN 3045 3063 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 862 864 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2779 2781 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2895 2897 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 2944 2944 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2947 2947 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2950 2950 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2959 2959 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2965 2965 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2968 2968 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2971 2971 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2983 2983 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3000 3000 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3003 3003 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3014 3014 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3023 3023 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3026 3026 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3029 3029 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 700 700 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 798 798 O-LINKED (XYL. .) (CHONDROITIN SULFATE)
(POTENTIAL).
FT CARBOHYD 889 889 O-LINKED (XYL. .) (CHONDROITIN SULFATE)
(POTENTIAL).
FT CARBOHYD 981 981 O-LINKED (XYL. .) (CHONDROITIN SULFATE)
(POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2206 2206 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2528 2528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 25 1188 Missing (in isoform Short).
FT FTid=VSP_001149.
SQ SEQUENCE 3063 AA; 333189 MW; 75FEA78FA8E48293 CRC64;
Query Match 25.0%; Score 255.5; DB 1; Length 3063;
Best Local Similarity 35.5%; Pred. No. 6.3e-13;
Matches 66; Conservative 40; Mismatches 63; Indels 17; Gaps 6;
QY 6 DIVILVDSGNSI--YPWVEVQHFLINILKKFYIGPGQIQGVGVQYGEDVVHFEPLNDYRS 63
Db 1199 DIVLLVDGWSIGRANPRTVRSFISRIVEFDIGPKRQVIALAQYSGDPTFETQLNAHRD 1258
QY 64 VKDVFAASHIEORGGTETATGIEPARSEAF--QKGRKGAKKVMIVITDGESHDSPD 121
Db 1259 KKSILQAVANLPYKGG-NTLTGMALNFIROQNFTQAGMRPRKIGVLITDGKSD--D 1315
QY 122 LKVKIQSREDNTRVAVVLGYNNRGINPETFELNEIKVADPDHKKFFNVITDEALK 191
Db 1316 VEAPSKKDKGEVLEFAI-----GIKNADDEV-----ELKMIATDPDPTHYNNVADFESLS 1365
QY 182 DIVDAL 187
```

Db 1366 RIVDDL 1371

RESULT 11

CALC_CHICK STANDARD; PRT; 3124 AA.

AC PI3944; Q04509;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(XII) chain precursor (Fibrochimerin).
GN COL12A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn;
RX MEDLINE=92011862; PubMed=1918137;
RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA Nishida Y., Obara M., Kimata K.;
RT "The complete primary structure of type XII collagen shows a chimeric
RT molecule with reiterated fibronectin type III motifs, von Willebrand
RT factor A motifs, a domain homologous to a noncollagenous region of
RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT site.";
RL J. Cell Biol. 115:209-221(1991).
RN [2]
RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP 2846-2873.
RX MEDLINE=90062079; PubMed=2584192;
RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
RT "Type XII collagen. A large multidomain molecule with partial
RT homology to type IX collagen.";
RL J. Biol. Chem. 264:19772-19778(1989).
RN [3]
RP SEQUENCE OF 2960-3076 FROM N.A.
RX MEDLINE=87317590; PubMed=3476925;
RA Gordon M.K., Gerecke D.R., Olsen B.R.;
RT "Type XII collagen: distinct extracellular matrix component
RT discovered by cDNA cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
RN [4]
RP SEQUENCE OF 1-1283 FROM N.A. (ISOFORM SHORT), AND ALTERNATIVE
RP SPLICING.
RC TISSUE=Embryo;
RX MEDLINE=93042014; PubMed=1420368;
RA Trueb J., Trueb B.;
RT "The two splice variants of collagen XII share a common 5' end.";
RL Biochim. Biophys. Acta 1171:97-98(1992).
RN [5]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95370352; PubMed=7642694;
RA Koch W., Bohrmann B., Matthison M., Hagios C., Trueb B., Chiquet M.;
RT "Large and small splice variants of collagen XII: differential
RT expression and ligand binding.";
RL J. Cell Biol. 130:1005-1014(1995).
CC -!- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the perifibrillar matrix.
CC -!- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=The final tissue form of collagen XII may contain
CC homotrimers of either isoform Long or isoform Short or any
CC combination of isoform Long and isoform Short. Only isoform Long
CC is a proteoglycan. Isoform Long has more restricted expression
CC in embryonic tissue than isoform Short;

```

CC Name=long;
CC IsoId=P13944-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P13944-2; Sequence=VSP_001148;
CC -!- TISSUE SPECIFICITY: Type XII collagen is present in tendons,
CC ligaments, perichondrium, and periosteum, all dense connective
CC tissues containing type I collagen.
CC -!- DOMAIN: This sequence defines five distinct domains, two triple-
CC helical domains (COL1 and COL2) and three nontriple-helical
CC domains (NC1, NC2, and NC3).
CC -!- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end.
CC -!- PTM: prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: O-glycosylated; glycosaminoglycan of chondroitin-sulfate type
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -!- SIMILARITY: Contains 4 WFPA domains.
CC -!- SIMILARITY: Contains 18 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00824; BAA00701.1; -
CC EMBL; X61024; CAA43358.1; -
CC EMBL; M17375; AAM48718.1; -
CC EMBL; J05137; AAM48635.1; -
CC EMBL; X67327; CAA47744.1; -
CC EMBL; PIR; A40020; A40020.
CC HSSP; P20701; 1LEA.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR008957; FN III-like.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR002035; WFPA.
CC Pfam; PF01391; Collagen; 4.
CC Pfam; PF00041; fn3; 17.
CC Pfam; PF02210; TSPN; 1.
CC Pfam; PF00092; vwa; 4.
CC PRINTS; PR00453; WVFADOMAIN.
CC SMART; SM00060; FN3; 16.
CC SMART; SM00210; TSPN; 1.
CC SMART; SM00327; WVA; 4.
CC PROSITE; PS50234; WVFA; 4.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
CC FT SIGNAL 1 24
CC FT CHAIN 25 3124
CC FT DOMAIN 25 114
CC FT DOMAIN 139 311
CC FT DOMAIN 332 425
CC FT DOMAIN 439 615
CC FT DOMAIN 629 720
CC FT DOMAIN 721 811
CC FT DOMAIN 812 904
CC FT DOMAIN 905 998
CC FT DOMAIN 999 1085
CC FT DOMAIN 1086 1178
CC FT DOMAIN 1199 1371
CC FT DOMAIN 1384 1473
CC FT DOMAIN 1474 1565
CC FT DOMAIN 1566 1654
CC FT DOMAIN 1655 1755
CC FT DOMAIN 1756 1846
CC FT DOMAIN 1847 1936
CC FT DOMAIN 1937 2027

```

```

FT DOMAIN 2028 2118
FT DOMAIN 2119 2206
FT DOMAIN 2207 2295
FT DOMAIN 2327 2500
FT DOMAIN 2524 2716
FT DOMAIN 2455 2750
FT DOMAIN 2751 2902
FT DOMAIN 2903 2945
FT DOMAIN 2946 3048
FT DOMAIN 3049 3124
FT DOMAIN 3086 3096
FT DOMAIN 3111 3123
FT DOMAIN 2438 2440
FT SITE 2899 2901
FT CARBOHYD 32 32
FT CARBOHYD 797 797
FT CARBOHYD 890 890
FT CARBOHYD 981 981
FT CARBOHYD 1006 1006
FT CARBOHYD 1032 1032
FT CARBOHYD 1044 1044
FT CARBOHYD 1512 1512
FT CARBOHYD 1767 1767
FT CARBOHYD 2210 2210
FT CARBOHYD 2273 2273
FT CARBOHYD 2532 2532
FT CARBOHYD 2683 2683
FT VARSPLIC 25 1188
FT CONFLICT 1258 1258
FT CONFLICT 1264 1264
FT CONFLICT 2759 2759
FT CONFLICT 2803 2803
FT CONFLICT 2977 2977
FT CONFLICT 3075 3076
FT CONFLICT 3124 AA; 340578 MW; 094285AFE7F346CF CRC64;
SQ SEQUENCE 3124 AA; 340578 MW; 094285AFE7F346CF CRC64;

Query Match 24.3%; Score 248.5; DB 1; Length 3124;
Best Local Similarity 34.4%; Pred. No. 2.4e-12;
Matches 66; Conservative 43; Mismatches 58; Indels 25; Gaps 8;

QY 1 CQT--YMDIVIVLDGNSI--YPMVEVQHFLINILKKFYIGPGQIQGVGVGYGVHEF 56
Db 1192 CRTAEADIVLLVDGWSIGRNFKTVRNFISRIVEFDIGDKVQIGLAQYSGDPRTEW 1251
QY 57 HLDYRSVKDVVEAASHIEQSGGTETRTAFGIEFARSEAFOK--GGRKGAKVMIVITDG 114
Db 1252 NLNAYRIKALDAVTNIPYKGG-NLTGMDALFILKNFKQEAGLRPRKIGVLIIDG 1310
QY 115 ESHDPSDLKVIQOSER---DNVTRYAVAVGYNNREGINPETFLNEIKYIASDDPKHF 171
Db 1311 KSQDD----VTFPSRRLRDEGVLYAIGIK-----MADE--NELKQIATDPPDIHA 1355
QY 172 ENVTDEAAAKDI 183
Db 1356 YNVADFSFLASI 1367

RESULT 12
ITAD MOUSE STANDARD; PRT; 1167 AA.
AC Q60677;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-E precursor (Integrin alpha M290).
GN ITGAE.
OS Mus musculus (Mouse).

```

FT	DOMAIN	185	191	GLU-RICH (ACIDIC).
FT	SITE	1140	1144	GFPR MOTIF.
FT	DISULFID	72	83	BY SIMILARITY.
FT	DISULFID	130	164	BY SIMILARITY.
FT	DISULFID	698	754	BY SIMILARITY.
FT	DISULFID	814	820	BY SIMILARITY.
FT	DISULFID	884	898	BY SIMILARITY.
FT	DISULFID	998	1023	BY SIMILARITY.
FT	DISULFID	1031	1047	BY SIMILARITY.
FT	CARBOHYD	51	51	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	256	256	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	314	314	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	341	341	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	364	364	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	418	418	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	718	718	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	773	773	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	829	829	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	846	846	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	925	925	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	968	968	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1013	1013	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1055	1055	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	1086	1086	N-LINKED (GLCNAC. . .) (POTENTIAL) .
SQ	SEQUENCE	1167 AA;	128983 MW;	B8331C115DCCCFD CRC64;

Query Match 24.0%; Score 245.5; DB 1; Length 1167;

Best Local Similarity 34.5%; Pred. No. 1.4e-12;

Matches 68; Conservative 34; Mismatches 84; Indels 11; Gaps 6

QY	6	DIVIVL	DGNSIYP--WVEVQH	FLINILKPYIGGQIQVG	VQYGVGV	QEDVWHE	FHLNDYRS	63
DB	195	EAIVL	DGSGISG	SPDKAKNFISTWMRNF	YKFCFCN	FALVQYGA	VIQTEPDLQ	254
QY	64	VKD	VEAASHLEQR	GGTETAFGL	IEFARSEAP--QKGRK	CAKVMIVITD	GSHDSP-	120
DB	255	INASL	KVQSIQVKEV-TKTAS	AMQHVLDNIFPSGR	SKRKALKVMVL	DGDFIGDPL	313	
QY	121	DLEK	VIQOSERD	NVTRVAVVLGYNNRR	GINPETFLEIK	YIASDPD	KHFENVTD	180
DB	314	NLTIV	NSPKMGVVR	FAIGV---GDRF	KNNNTY-RELK	LIASDPKEA	HTFKVTNISAL	368
QY	181	KOIVD	ALGDRIFS	LEGT	197			
DB	369	DGLLS	KLQQRIVHMEGT	385				

RESULT 13

ITAE_HUMAN STANDARD; PRT; 1179 AA.

AC P38570; Q9NZU9;

DT 01-OCT-1994 (Rel. 30, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Integrin alpha-E precursor (Mucosal lymphocyte-1 antigen) (HML-1 antigen) (CD103 antigen) (Integrin alpha-IEL).

GN ITGAE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-38 AND 179-188.

RC TISSUE=Leukemia, and Lymphocytes;

RX MEDLINE=94164962; PubMed=8119947;

RA Shaw S.K., Cepek K.L., Murphy E.A., Russell G.J., Brenner M.B., Parker C.M.;

RT "Molecular cloning of the human mucosal lymphocyte integrin alpha E subunit. Unusual structure and restricted RNA distribution.";

RL J. Biol. Chem. 269:6016-6025(1994)

EN [2]

DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha
 DE subunit) (CR-3 alpha chain) (CD11b) (leukocyte adhesion receptor M01).
 GN ITGM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88312584; PubMed=3044779;
 RA Pytela R.;
 RA "Amino acid sequence of the murine Mac-1 alpha chain reveals homology
 RT with the integrin family and an additional domain related to von
 RT Willebrand factor";
 RL EMO J. 7:1371-1378 (1988).
 RN [2]
 RP SEQUENCE OF 11-45 FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Spleen;
 RX MEDLINE=86287312; PubMed=2942940;
 RA Sastre L., Roman J.M., Teplow D.B., Dreyer W.J., Gee C.E.,
 RA Larson R.S., Roberts T.M., Springer T.A.;
 RT "A partial genomic DNA clone for the alpha subunit of the mouse
 RT complement receptor type 3 and cellular adhesion molecule Mac-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5644-5648 (1986).
 RN [3]
 RP SEQUENCE OF 17-28.
 RX MEDLINE=8518276; PubMed=3887182;
 RA Springer T.A., Teplow D.B., Dreyer W.J.;
 RT "Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion
 RT glycoproteins and unexpected relation to leukocyte interferon.";
 RL Nature 314:540-542 (1985).
 CC -!- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS
 CC ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES
 CC AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.
 CC IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF
 CC THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D
 CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR
 CC FIBRINOGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES
 CC OF FIBRINOGEN GAMMA CHAIN. ALPHA-M/BETA-2 PLAY A CRITICAL ROLE IN
 CC MAST CELL DEVELOPMENT AND IN IMMUNE COMPLEX-MEDIATED
 CC GLOMERULONEPHRITIS. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-M
 CC SUBUNIT GENE DEMONSTRATE INCREASE IN NEUTROPHIL ACCUMULATION, IN
 CC RESPONSE TO A IMPAIRED DEGRANULATION AND PHAGOCYTOSIS, EVENTS THAT
 CC APPARENTLY ACCELERATE APOPTOSIS IN NEUTROPHILS. THESE MICE DEVELOP
 CC OBESITY.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN. ALPHA-M
 CC ASSOCIATES WITH BETA-2.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND
 CC GRANULOCYTES.
 CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
 CC -!- SIMILARITY: Contains 1 VFMA domain.
 CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X07640; CAA30479.1; -;
 CC EMBL; M14293; AAA39484.1; -;
 CC PIR; S00551; S00551.
 CC HSSP; P11215; IABX.
 CC MGD; MGI:96607; Itgam.
 CC InterPro; IPR00413; Integrin_alpha.
 CC InterPro; IPR002035; VFMA_A.

DR Pfam; PF01839; FG-GAP; 3.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA_1
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS0234; VVFA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Calcium; Repeat.
 FT SIGNAL 1 16
 FT CHAIN 17 1153 INTEGRIN ALPHA-M.
 FT DOMAIN 17 1105 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1106 1129 POTENTIAL.
 FT DOMAIN 1130 1153 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 31 84 FG-GAP 1.
 FT REPEAT ? ? FG-GAP 2.
 FT DOMAIN 164 350 VVFA.
 FT REPEAT 337 400 FG-GAP 3.
 FT REPEAT 401 452 FG-GAP 4.
 FT REPEAT 454 515 FG-GAP 5.
 FT REPEAT 517 575 FG-GAP 6.
 FT REPEAT 580 632 FG-GAP 7.
 FT CA_BIND 465 473 POTENTIAL.
 FT CA_BIND 529 537 POTENTIAL.
 FT CA_BIND 592 600 POTENTIAL.
 FT SITE 1132 1136 GFFKR MOTIF.
 FT DISULFID 66 73 BY SIMILARITY.
 FT DISULFID 105 123 BY SIMILARITY.
 FT DISULFID 654 711 BY SIMILARITY.
 FT DISULFID 770 776 BY SIMILARITY.
 FT DISULFID 999 1023 BY SIMILARITY.
 FT DISULFID 1028 1033 BY SIMILARITY.
 FT CARBOHYD 58 58 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 86 86 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 391 391 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 696 696 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 734 734 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 772 772 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 801 801 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 881 881 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 907 907 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 941 941 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 980 980 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 994 994 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1022 1022 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1045 1045 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1051 1051 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1076 1076 N-LINKED (GLCNAC.) (POTENTIAL).
 SQ SEQUENCE 1153 AA; 127480 MW; 178DB988AECB0343 CRC64;

Query Match 23.5%; Score 240.5; DB 1; Length 1153;
 Best Local Similarity 32.8%; Pred. No. 3.5e-12;
 Matches 66; Conservative 41; Mismatches 81; Indels 13; Gaps 7;
 QY 2 QTYMDIVILVDGNSI--YPWVEVQHFLINILKFKYIGPQIQGVVQYGVGVVHEPLN 59
 Db 146 QBSDIVFLIDGSGSINNIDFQWKKEFVSVVMQF--KSKTLFLSLMQYSDERHFTFN 203
 QY 60 DYBSKDVVEAASHIEQGGTETRTAFGEFARSEAFQK--GGRKAKKVMIVITGESH 117
 Db 204 DFKRNPSPRSHVSPKQLNG-RTKTAGSRKVVVRELPHKTNAGARENAKILVITDGEKF 262
 QY 118 DSP-DLEKTIQOSQONVTRYAVAVLGYNNRGINPETFNLNEIKYTASDPDDKHFNVD 176
 Db 263 GDPDLVDKVIDPADRAGVIRYVIGVNAFNK----PQS-RRELDTIASKPAGEHVQVDN 317
 QY 177 EALKDIVDALGDRIFSLEGT 197
 Db 318 FEALNTIQQLQEKIFAIEGT 338

```

RESULT 15
CAIC NOTVI
ID CAIC NOTVI STANDARD; PRT; 929 AA.
AC Q91145;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(XII) chain (Fragment).
OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
OC Notophthalmus.
OC NCBI_TaxID=8316;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=95246925; PubMed=7729595;
RX Wei Y., Yang E.V., Klatt K.P., Tassava R.A.;
RA "Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type
RT XII collagen, a developmentally regulated extracellular matrix
RT protein in regenerating newt limbs."
RL Dev. Biol. 168:503-513(1995).
CC -!- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the pericellular matrix (By similarity). Could play
CC a developmental role in regeneration.
CC -!- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences (By similarity).
CC -!- DEVELOPMENTAL STAGE: Expression starts at 3 days after amputation
CC in cells of the basal layer of the wound epithelium. At day 10,
CC expression is found in both the basal wound epithelial cells and
CC the distal mesenchyme cells. At mid-bud and late-bud blastema
CC stages, wound epithelium expression has decreased, whereas the
CC mesenchyme remains strongly active in transcription and showed a
CC tendency toward distal regionalization. Condensing cartilage shows
CC no signal. Finally, at the late digit stage, expression becomes
CC largely restricted to the perichondrium.
CC -!- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end (By similarity).
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: Contains 2 VWFA domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U19494; AAA80217.1; -.
CC PIR; I51027; I51027.
CC HSP; P02751; LFNA.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR002035; VWFA.
CC Pfam; PF00041; fn3; 7.
CC Pfam; PF00092; vwa; 2.
CC PRINTS; PR00453; VWFADOMAIN.
CC SMART; SM00060; FN3; 4.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS02334; VWFA; 2.
CC EXTRACELLULAR MATRIX; Connective tissue; Repeat; Cell adhesion;
KW Collagen; Glycoprotein.
FT NON_TER 1
FT DOMAIN <1 49 VWFA 1.
FT DOMAIN 63 154 FIBRONECTIN TYPE-III 1.
FT DOMAIN 155 245 FIBRONECTIN TYPE-III 2.
FT DOMAIN 246 338 FIBRONECTIN TYPE-III 3.

```

```

FT DOMAIN 339 432 FIBRONECTIN TYPE-III 4.
FT DOMAIN 433 519 FIBRONECTIN TYPE-III 5.
FT DOMAIN 520 612 FIBRONECTIN TYPE-III 6.
FT DOMAIN 633 805 VWFA 2.
FT DOMAIN 818 907 FIBRONECTIN TYPE-III 7.
FT DOMAIN 908 >929 FIBRONECTIN TYPE-III 8.
FT CARBOHYD 231 231 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
(POTENTIAL).
FT CARBOHYD 324 324 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
(POTENTIAL).
FT CARBOHYD 415 415 O-LINKED (XYL. . .) (CHONDROITIN SULFATE)
(POTENTIAL).
FT CARBOHYD 98 98 CARBOHYD
FT NON_TER 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 929 AA; 101647 MW; AE5D7485254FD954 CRC64;

Query Match 23.2%; Score 237.5; DB 1; Length 929;
Best Local Similarity 31.6%; Pred. No. 4.8e-12;
Matches 61; Conservative 49; Mismatches 60; Indels 23; Gaps 7;

QY 6 DIVIVLDGNSI--YPWVEVQHFLINILKKFYIGPGQIQVGVVOYGEDVHVHEPHLNDYRS 63
DB 633 DIVILLVDSGWSIGRPNKIVNFISRVVEVDIGSDRVQIVASQYSGDPRTQWLNTHKT 692
QY 64 VKDVVEAASHIEQRGGTETRTAFGIEFARSAFOK--GGRKGAKKVMIVITDGHSHSPD 121
DB 693 KKSIMDAVANILPYKGG-NTNTGSALKFLENNFRPGVGMREKARKIALLLTDGKSQDD-- 749
QY 122 LEKVIQOSER---DNVTRYAVAVLGVYNNRGINETFLNEIKYIASDPDDKKHFNVTDEA 178
DB 750 ---IVAPSKRYADEGIELYAVGIK-----NADE--NELKEIASDPDELYMYNVADFS 796
QY 179 ALKDIVDALGDRI 191
DB 797 LLTNIVNDLTENV 809

Search completed: September 21, 2004, 13:00:05
Job time : 7.77793 secs

```

Result No.	Query	Score	Match		Length	DB	ID	Description
			%					
1	1010	98.8	1188	11	Q7TC03	Q7tc03 mus musculus		
2	598	58.5	1171	13	O42094	O42094 gallus gall		
3	408	39.9	191	6	Q29124	Q29124 sus scrofa		
4	365	35.7	323	11	Q8C584	Q8cb84 mus musculus		
5	254.5	24.9	1160	6	Q8MK54	Q8mkf4 felis silve		
6	245.5	24.0	895	11	Q9WUF8	Q9wfuf8 mus sp. itg		
7	245.5	24.0	1167	11	O88341	O88341 rattus norv		
8	245.5	24.0	1167	11	O88340	O88340 rattus norv		
9	241	23.6	957	4	Q9H0V3	Q9h0v3 homo sapien		
10	240.5	23.5	1036	11	O8CA73	O8ca73 mus musculus		
11	240.5	23.5	1038	11	O8BS01	O8be01 mus musculus		
12	240	23.5	954	4	Q8WXY8	Q8wxv8 homo sapien		
13	240	23.5	957	4	Q36P44	Q96p44 homo sapien		
14	235.5	23.0	1797	11	O80X19	O80x19 mus musculus		
15	225	22.0	517	4	O43853	O43853 homo sapien		
16	223.5	21.9	1151	11	O9J130	O9ji30 rattus norv		

and

219 YPSVKNVFEAAASHIEPBCCTETPTAT ECTEEFAACF

n^b

219 YPSVKNDVLEAASCHIEPABCCCTETPTAEPCTEFPAPCCEAFQYGCGRK708

```

Qy 121 DLEKVIQOBSRDNVTRAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL 180
Db 279 DLEKVIQSEKDNVTRAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAAL 338

Qy 181 KOIVDALGDRIFSLEGT 197
Db 339 KOIVDALGDRIFSLEGT 355

RESULT 2
O42094 PRELIMINARY; PRT; 1171 AA.
ID O42094
AC O42094;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ALPHAI integrin
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gizzard;
RX MEDLINE=97476270; PubMed=9334246;
RA Obata H., Hayashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,
RA Sobue K.;
RT "Smooth muscle cell phenotype-dependent transcriptional regulation of
RT the alphas integrin gene.";
RL J. Biol. Chem. 272:26643-26651(1997).
DR EMBL; AB000470; BAA23160.1; -.
DR EMBL; AB000471; BAA23161.1; -.
DR PIR; A55348; A55348.
DR HSSP; P17301; IAOX.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; EG-GAP; 3.
DR Pfam; PF0357; integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWF_A; 1.
KW Integrin.
SQ SEQUENCE 1171 AA; 130228 MW; B505A4D65F09736E CRC64;

Query Match 58.5%; Score 598; DB 13; Length 1171;
Best Local Similarity 57.3%; Pred. No. 3.1e-44;
Matches 114; Conservative 39; Mismatches 44; Indels 2; Gaps 1;

Qy 1 CQYMDIVIVLDGNSIYPWVEVQHFLINILKFKYIGPGQIQGVVQYGEDVWVHFLND 60
Db 157 CKTQLDIVIVLDGNSIYPWVSTAFNLNLRMDIGPQQTQGVIGVQGTWVHFLNT 216

Qy 61 YRSVKDVVAASHIEORGTEHTAFGIEFAFSEAFQK--GGKRGAKKVMIVTDGESH 118
Db 217 YSTTEEVDAALRIORGTEHTAFGIEFAFSEAFQK--GGKRGAKKVMIVTDGESH 276

Qy 119 SPDEKVIQOBSRDNVTRAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTD 178
Db 277 NVRLQEVDCEDENTQRFALILGYSRGNLSTEFVVEIKSIASKTEKHFNFVSD 336

Qy 179 ALKDIVDALGDRIFSLEGT 197
Db 337 ALTVIVEALGERIFALEAT 355

```

RESULT 3

```

Q29124 PRELIMINARY; PRT; 191 AA.
ID Q29124
AC Q29124;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE VLA-2 (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=95036279; PubMed=7949129;
RA Babou W.F., Potter C.L., Mirza H.;
RA "The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific
RT recognition sequence for endothelial cell attachment and spreading;
RT molecular and functional characterization.";
RL Blood 84:3734-3741(1994).
DR EMBL; Z12137; CAA78125.1; -.
DR PIR; I47230; I47230.
DR HSSP; P17301; IAOX.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWF_A; 1.
FT NON_TER 1 1
FT NON_TER 191 191
SQ SEQUENCE 191 AA; 20896 MW; 2E228B472EC699D8 CRC64;

Query Match 39.9%; Score 408; DB 6; Length 191;
Best Local Similarity 45.6%; Pred. No. 2.1e-28;
Matches 82; Conservative 35; Mismatches 61; Indels 2; Gaps 1;

Qy 20 WVEVQHFLINILKFKYIGPGQIQGVVQYGEDVWVHFLNDYRSVKDVVAASHIEQ 79
Db 1 WDAVKNFLEKVFQGLDIGTQTKQGLIQYANNPRVFNITFKAEWVEATSHHTQYGG 60

Qy 80 TETRTAFGIEFAFSEAFQ--KGGKRGAKKVMIVTDGESHDPDLLEKVIQOBSRDNV 137
Db 61 DLTNTEKAIQYARDSAYSAAAGRGFGATKMWVVVTDGESHGSMUKAVIDQCNDN 120

Qy 138 AVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDAAALKDIVDALGDRIFS 197
Db 121 GIAVLGYLNRLNADTKNLKIKETASITPTERYFFNVSDADLLEKAGTLGQIF 180

RESULT 4
Q8CB84 PRELIMINARY; PRT; 323 AA.
ID Q8CB84
AC Q8CB84;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Integrin alpha 2 (Fragment).
GN ITGA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";

```

Matches	69;	Conservative	30;	Mismatches	85;	Indels	13;	Gaps	5;
QY	6	DIVILDGNSIYP--WVEVQHFLINILTKFYIGPGQIQGVVVOYGEDVVHFLINDYRS	63						
Db	191	EAIIILDGSGIDPPDFQAKQDFISNMKNFYAKCFECFNALVOYGEVICTEFLDRSQD	250						
QY	64	VKDVEAASHIEQRGTETRTAFGIFBARSEAF--QKGGKKGAKVMIVITTDGESHDSP-	120						
Db	251	ALASLARVQNIQTQKNV-IKTASAIQHVLDNIITPPSHGSRKNAPKIVIVITDGTGDPL	309						
QY	121	DLEKVIQCSERDNTVYAVAVLGYNRRGINPTFTNEIKYIASDDDDKHFFNVTVDEAAL	180						
Db	310	NLTVTISSPQMQVERFAIRV-----GNESKTKLKLKLIASDPFERHAFTVTNYSAL	362						
QY	181	KDIVDALGDRIFSLGFT	197						
Db	363	DGLLSKLQNIHTGFT	379						
RESULT	6								
Q9WUF8									
ID	Q9WUF8	PRELIMINARY;	PRT;	895	AA.				
AC	Q9WUF8;								
DT	01-NOV-1999	(TEMBLrel. 12, Created)							
DT	01-NOV-1999	(TEMBLrel. 12, Last sequence update)							
DT	01-JUN-2003	(TEMBLrel. 24, Last annotation update)							
DE	Itgae protein (Fragment)								
GN	ITGAE.								
OS	Mus sp.								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Rodentia; Sciurognathii; World								

```

RN      SEQUENCE FROM N.A.
RP
RA      Schoen M.P., Aya A., Murphy E.A., Adams C.M., Strauch U.G.,
RA      Agace W.W., Marsal J., Donohue J.P., Her H., Beier D.R., Olson S.,
RA      Lefrancois L., Brenner M.B., Grusby M.J., Parker C.M.;
RT      "Mucosal T lymphocyte numbers are selectively reduced in integrin
RT      alphaE (CD103) deficient mice.";
RL      Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF133085; AAD30063.1; -.
DR      EMBL; AF133070; AAD30063.1; JOINED.
DR      EMBL; AF133071; AAD30063.1; JOINED.
DR      EMBL; AF133072; AAD30063.1; JOINED.
DR      EMBL; AF133073; AAD30063.1; JOINED.
DR      EMBL; AF133074; AAD30063.1; JOINED.
DR      EMBL; AF133075; AAD30063.1; JOINED.
DR      EMBL; AF133076; AAD30063.1; JOINED.
DR      EMBL; AF133077; AAD30063.1; JOINED.
DR      EMBL; AF133078; AAD30063.1; JOINED.
DR      EMBL; AF133079; AAD30063.1; JOINED.
DR      EMBL; AF133080; AAD30063.1; JOINED.
DR      EMBL; AF133081; AAD30063.1; JOINED.
DR      EMBL; AF133082; AAD30063.1; JOINED.
DR      EMBL; AF133083; AAD30063.1; JOINED.
DR      EMBL; AF133084; AAD30063.1; JOINED.
DR      HSP; P11215; 1BHQ
DR      GO; GO:008305; C:integrin complex; IEA.
DR      GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR      GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR      InterPro; IPR000413; Integrin_alpha.
DR      InterPro; IPR002035; VWF_A.
DR      Pfam; PF01839; FG-GAP; 3.
DR      Pfam; PF00092; vwa; 1.
DR      PRINTS; P01185; INTEGRINA.
DR      PRINTS; P00453; VWFADOMAIN.
DR      SMART; SM00191; Int_alpha; 4.
DR      SMART; SM00327; VWA; 1.
DR      PROSITE; PSS0234; VWF_A; 1.
DR      NON_TER      1
DR      FT      1
DR      NON_TER      895
DR      SEQUENCE      895 AA; 98266 MW; BE6EA14A754DA1E CRC64;
SQ

```

Query Match 24.0%; Score 245.5; DB 11; Length 895;
Best Local Similarity 34.0%; Pred. No. 5.6e-13;
Matches 67; Conservative 35; Mismatches 84; Indels 11; Gaps 6;

QY 6 DIVVLGSGNSIYP--WVEVQHFLINILKKFYIGPGQIQGVVQYGVGVVHFEHLNDYRS 63
DB 182 EIAIVLDGSGSIEPSDFQAKNFISTMWRNFEKFCFNCFNALVQYGVAVIQTEFDLQESRD 241

QY 64 VKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHSP- 120
DB 242 INASLAKVQSIQVQKEV-TKTASAMQHVLNIFIPSRGSRKKALKVMVLTGDIQFGDPL 300

QY 121 DLEKVIQOSERDNTVRVAVLGVYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180
DB 301 NLTTVINSPKMGVVRFAIGVDAFK-----NNNTY-RELKLIASDPKHAHTFKVTNYSAL 355

QY 181 KDIVDALGDRIFSLEGT 197
DB 356 DGLLSKLOQRIVHMEGT 372

RESULT 7
O88341 ID O88341 PRELIMINARY; PRT; 1167 AA.
AC O88341;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Integrin alpha E2 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98056820; PubMed=9394838;
RA Brenan M., Rees D.J.G.;
RT "Sequence analysis of rat integrin alpha E1 and alpha E2 subunits:
RT tissue expression reveals phenotypic similarities between
RT intraepithelial lymphocytes and dendritic cells in lymph."
RL Eur. J. Immunol. 27:3070-3079(1997).
DR EMBL; AF020046; AAC23663.1; -
DR HSSP; P11215; 1BHQ.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; EG-GAP; 3.
DR Pfam; PF00357; Integrin_A; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR NON_TER 1
SQ SEQUENCE 1167 AA; 128593 MW; E3FED7E18B5CBBEC CRC64;

Query Match 24.0%; Score 245.5; DB 11; Length 1167;
Best Local Similarity 35.0%; Pred. No. 8.1e-13;
Matches 69; Conservative 32; Mismatches 85; Indels 11; Gaps 6;

QY 6 DIVVLGSGNSIYP--WVEVQHFLINILKKFYIGPGQIQGVVQYGVGVVHFEHLNDYRS 63
DB 195 EIAIVLDGSGSIEPSDFQAKNFISTMWRNFEKFCFNCFNALVQYGVAVIQTEFDLQESRD 254

QY 64 VKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHSP- 120
DB 255 INASLAKVQSIQVQKEV-TKTASAMQHVLNIFIPSRGSRKKALKVMVLTGDIQFGDPL 313

QY 121 DLEKVIQOSERDNTVRVAVLGVYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180
DB 313 NLTTVINSPKMGVVRFAIGVDAFK-----NNNTY-RELKLIASDPKHAHTFKVTNYSAL 368

QY 181 KDIVDALGDRIFSLEGT 197
DB 369 DGLLSKLOQRIVHMEGT 385

RESULT 9
O88343 ID O88343 PRELIMINARY; PRT; 957 AA.
AC O88343;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Integrin alpha E1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98056820; PubMed=9394838;
RA Brenan M., Rees D.J.G.;
RT "Sequence analysis of rat integrin alpha E1 and alpha E2 subunits:
RT tissue expression reveals phenotypic similarities between
RT intraepithelial lymphocytes and dendritic cells in lymph."
RL Eur. J. Immunol. 27:3070-3079(1997).
DR EMBL; AF020045; AAC23662.1; -
DR HSSP; P11215; 1BHQ.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; EG-GAP; 3.
DR Pfam; PF00357; Integrin_A; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR NON_TER 1
SQ SEQUENCE 1167 AA; 128970 MW; D88A2C38ACDC2AAE CRC64;

Db 314 NLTTVISSSKMQGVVRFAIGVNAFE-----NNNTY-RELKLIASDPKHAHTFKVTNYSAL 368

QY 181 KDIVDALGDRIFSLEGT 197

Db 369 DGLLSKLOQRIVHMEGT 385

RESULT 8
O88340 ID O88340 PRELIMINARY; PRT; 1167 AA.
AC O88340;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Integrin alpha E1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98056820; PubMed=9394838;
RA Brenan M., Rees D.J.G.;
RT "Sequence analysis of rat integrin alpha E1 and alpha E2 subunits:
RT tissue expression reveals phenotypic similarities between
RT intraepithelial lymphocytes and dendritic cells in lymph."
RL Eur. J. Immunol. 27:3070-3079(1997).
DR EMBL; AF020045; AAC23662.1; -
DR HSSP; P11215; 1BHQ.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; EG-GAP; 3.
DR Pfam; PF00357; Integrin_A; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR NON_TER 1
SQ SEQUENCE 1167 AA; 128970 MW; D88A2C38ACDC2AAE CRC64;

Query Match 24.0%; Score 245.5; DB 11; Length 1167;
Best Local Similarity 34.0%; Pred. No. 8.1e-13;
Matches 67; Conservative 35; Mismatches 84; Indels 11; Gaps 6;

QY 6 DIVVLGSGNSIYP--WVEVQHFLINILKKFYIGPGQIQGVVQYGVGVVHFEHLNDYRS 63
DB 195 EIAIVLDGSGSIEPSDFQAKNFISTMWRNFEKFCFNCFNALVQYGVAVIQTEFDLQESRD 254

QY 64 VKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDGESHSP- 120
DB 255 INASLAKVQSIQVQKEV-TKTASAMQHVLNIFIPSRGSRKKALKVMVLTGDIQFGDPL 313

QY 121 DLEKVIQOSERDNTVRVAVLGVYNNRRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180
DB 314 NLTTVINSPKMGVVRFAIGVDAFK-----NNNTY-RELKLIASDPKHAHTFKVTNYSAL 368

QY 181 KDIVDALGDRIFSLEGT 197
DB 369 DGLLSKLOQRIVHMEGT 385

RESULT 9
O88343 ID O88343 PRELIMINARY; PRT; 957 AA.
AC O88343;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Integrin alpha E1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98056820; PubMed=9394838;
RA Brenan M., Rees D.J.G.;
RT "Sequence analysis of rat integrin alpha E1 and alpha E2 subunits:
RT tissue expression reveals phenotypic similarities between
RT intraepithelial lymphocytes and dendritic cells in lymph."
RL Eur. J. Immunol. 27:3070-3079(1997).
DR EMBL; AF020045; AAC23662.1; -
DR HSSP; P11215; 1BHQ.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; EG-GAP; 3.
DR Pfam; PF00357; Integrin_A; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR NON_TER 1
SQ SEQUENCE 1167 AA; 128970 MW; D88A2C38ACDC2AAE CRC64;

```

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN DKFZP564B052.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansong W., Boecker M., Bloecher H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.,
RA "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL Genome Res. 11:422-435(2001).
DR EMBL; AL136624; CAB66559.1; -.
DR HSSP; P17301; 1A0X.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0007155; F:cell adhesion; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR002035; VWF A.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR ProDom; PD000007; Clg helix; 4.
DR SMART; SMC0327; VWA; 1.
DR SMART; SMC0210; TSPN; 1.
DR PROSITE; PS50234; VWF A; 1.
KW Hypothetical protein; Collagen.
SQ SEQUENCE 957 AA; 99413 MW; 391022F715024571 CRC64;

Query Match 23.6%; Score 241; DB 4; Length 957;
Best Local Similarity 34.5%; Pred. No. 1.5e-12;
Matches 67; Conservative 33; Mismatches 78; Indels 16; Gaps 6;

QY 1 CQTY-MDIVLVDGNSIYP--WVEVQHFLINILKFFYIGPGQIQGVGVQVGEDVWHEFH 57
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 31 CRTAPTDLVFLDGSYSGVPENFEIVKKLVNITKDFDIPKFTIQGVGVQVSDYFVLEIP 90
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 LNDYRSKDVVVEAASHIEQGGTETRTAFGIEFARSEAFQKGGKKGAKKVMIVITDGESH 117
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 91 LGSYDSGEHLTAAVESILYLG--NTKTGKAIQFALDYLFDKSSR-FLTKIAVLITDGK 148
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 DSPLEKVIQGSERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTD 177
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 149 D--DVKDAQAARDSKITLFAI-----GVGSETDELRAANKPSSTVYFYVEDY 197
QY 178 AALKDVIDDALGDRI 191
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 198 IATSKIRENVKQKL 211

RESULT 10
Q8CAV73 ID Q8CAV73 PRELIMINARY; PRT; 1036 AA.
AC Q8CAV73
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Integrin alpha M.
GN ITGAM OR F730045J24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spinal cord;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK039444; BAC30350.1; -.
DR PIR; PT0572; PT0572.
DR PIR; PT0697; PT0697.
DR MGD; MGI:196607; Itgam.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:004895; F:cell adhesion receptor activity; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF A.
DR Pfam; PF01833; FG-GAP; 1.
DR Pfam; PF00357; integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SMC0327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWF A; 1.
SQ SEQUENCE 1036 AA; 115126 MW; 4F4F39BF1E188E77 CRC64;

Query Match 23.5%; Score 240.5; DB 11; Length 1036;
Best Local Similarity 32.8%; Pred. No. 1.9e-12;
Matches 66; Conservative 41; Mismatches 81; Indels 13; Gaps 7;

QY 2 QTYMDIVLVDGNSI--YPWVEVQHFLINILKFFYIGPGQIQGVGVQVGEDVWHEFLN 59
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 146 QESDIVFLDGGSGINNIDFQMKKEFVSTVMEQF--KSKTLFSLMQYSDERFHFTEFN 203
QY 60 DYRSKDVVVEAASHIEQGGTETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGESH 117
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 204 DFKRNPSPRHVSPIKOLNG-RITKASGIRKVVRELPHKTNAGARENAAKILVITDGEKF 262
QY 118 DSP--DLEKVIQGSERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFFNVTD 176
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 263 GDPIDYKDVPEADRAGVIRYVIGVGNVFNK---FQS--RELDITASKPAGEHVFQVDN 317
QY 177 EALKDVIDDALGDRIFSLEGT 197
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 318 FEALNTIQNLQEKIFAIEGT 338

RESULT 11
Q8BS01 ID Q8BS01 PRELIMINARY; PRT; 1038 AA.
AC Q8BS01
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Integrin.
GN ITGAE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK040983; BAC30769.1; -.
DR MGD; MGI:1298377; Itgae.

```


Db 31 CRTAPDLVFLDGSYSVCPENFEIVKKWLNVNITKNFDIGPKFIQGVVQVSDYVPLEIP 90
 Qy 58 LNDYRSKDVVEAASHIEQGGTETRTAFGIEFARSEAFQGRKGAKKVMIVITDGESH 117
 Db 91 LGSVDGSHLTAVESILYLG-NTKTGKALCFALDYLFAKSSR-FTTKIAVILTDGKSQ 148
 Qy 118 DSPLEKVIQOSRDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVITDE 177
 Db 149 D--DVKQAAQARDSKITLFAI-----GVSEIEDAELRAIANKPSSTYVYVEDY 197
 Qy 178 AALKDIDVALGDRI 191
 Db 198 IAIKIREWMKQKL 211

RESULT 14

Q80X19 PRELIMINARY; PRT; 1797 AA.
 AC Q80X19;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Collagen type XIV precursor.
 GN COL14A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1;
 RA Gerecke D.R., Meng X., Bin L., Birk D.E.;
 RT "Complete primary structure and genomic organization of the mouse
 coll14a1 gene."; 0:0-0(2003).
 RL Matrix Biol. 0:0-0(2003).
 DR EMBL; AY221110; AA064442.1; -.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR003962; Follin_subd.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01391; Collagen; 3.
 DR Pfam; PF00041; fn3; 8.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00092; vwa; 2.
 DR PRINTS; PR00014; FNTPPIII.
 DR SMART; SM00060; FN3; 8.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 2.
 DR PROSITE; PS02034; VWFA; 2.
 KW Collagen; Signal.
 FT SIGNAL
 SQ SEQUENCE 1797 AA; 193114 MW; 4109D5C4F6F78FA CRC64;

Query Match 23.0%; Score 235.5; DB 11; Length 1797;
 Best Local Similarity 31.1%; Pred. No. 1.1e-11;
 Matches 60; Conservative 42; Mismatches 72; Indels 19; Gaps 6;
 Qy 1 CQT--YMDIVLVDGNSI--YPWVEVQHFLINILKKFVIGPGIQGVVQVGEDVVEHF 56
 Db 152 CETPAIDIVLVDGNSIGRFNRLNLENVLTAFNVGSEKTRIGLAQYSGDPRIEW 211
 Qy 57 HLNDYRSKDVVEAASHIEQGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITDG 114
 Db 212 HLNAFTKDEVIDAVSLPYKGG-NTLTGLALNFENSPKPEAGSGVSKIGILITDG 270
 Qy 115 ESHDSPLEKVIQOSRDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNV 174
 Db 271 KSQD--DIIPSRNLRESGVLEFAIGV-----KNADLSLQELIAEPDSTHYNV 318

Qy 175 TDEAALKDIDVAL 187
 Db 319 AEFDMHTVVEST 331
 RESULT 15
 Q43853 PRELIMINARY; PRT; 517 AA.
 ID Q43853
 AC Q43853;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE TYPE XII collagen (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cornea;
 RA Wessel H., Anderson S., Fite D., Halvas E., Hempel J., SundarRaj N.;
 RT "Type XII collagen contributes to diversities in human corneal and
 limbal extracellular matrices."; 1
 RL Invest. Ophthalmol. Vis. Sci. 38:2408-2422 (1997).
 DR EMBL; U68139; AAC01506.1; -.
 DR HSSP; P17301; IAOX.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00041; fn3; 3.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 3.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS02034; VWFA; 1.
 KW Collagen.
 FT NON TER 1
 FT NON TER 517
 SQ SEQUENCE 517 AA; 56727 MW; 9B6972F44A1BD88F CRC64;
 Query Match 22.0%; Score 225; DB 4; Length 517;
 Best Local Similarity 33.2%; Pred. No. 1.8e-11;
 Matches 64; Conservative 35; Mismatches 68; Indels 26; Gaps 7;
 Qy 1 CQTYMDI---VIVLDGNS--IYPWVEVQHFLINILKKFVIGPGIQGVVQVGEDVVEHF 55
 Db 133 CSRGVDIKADMFLVDGSGYSIGIANFVKVRAFLVILVKSFEISPNRQISLVQYSRDPHTE 192
 Qy 56 FHLDYRSKDVVEAASHIEQGGTETRTAFGIEFARSEAF--QKGRKGAKKVMIVITD 113
 Db 193 FTLKFTKVEDIIEAINTFPYRGGS-TNTGKAMTYREKIFVPSKGRSNVVPKWILITD 251
 Qy 114 GESHDS---PDLEKVIQOSRDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKH 170
 Db 252 GKSSDAFRDPAI-----KLNRSDEIFAVGV-----KQAVDSELAIASPPAETH 296
 Qy 171 FPNVTDEAALKDI 183
 Db 297 VFTVEDFADFQRI 309

Search completed: September 21, 2004, 13:03:51
 Job time : 32.7151 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 13:03:59 ; Search time 35.493 Seconds
(without alignments)
1782.414 Million cell updates/sec

Title: US-09-980-403-2_COPY_159_355

Perfect score: 1022

Sequence: 1 CQYMDIVILVLDGNSIYPW.....AALKDIVDALGDRIFSLEGT 197

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1022	100.0	707	9	US-09-764-870-313
2	1022	100.0	707	14	US-10-125-540-313
3	1022	100.0	1034	10	US-09-984-130-43
4	1022	100.0	1034	10	US-09-836-353A-43
5	1022	100.0	1120	12	US-10-262-839-6
6	1022	100.0	1188	15	US-10-291-265-338
7	1022	100.0	1188	15	US-10-291-265-810
8	1022	100.0	1189	10	US-09-984-130-35
9	1022	100.0	1189	10	US-09-836-353A-35
10	1022	100.0	1189	12	US-10-262-839-4
11	1018	99.6	437	15	US-10-108-260A-3386
12	1002	98.0	193	10	US-09-805-354-8
13	1002	98.0	193	12	US-09-758-493-8
14	1002	98.0	193	14	US-10-144-259-8
15	599	58.6	1151	10	US-09-984-130-103

16	599	58.6	1151	10	US-09-836-353A-103	Sequence 103, App
17	599	58.6	1179	12	US-09-918-715-250	Sequence 250, App
18	599	58.6	1180	12	US-09-918-715-307	Sequence 307, App
19	596	58.3	212	9	US-09-996-738-6	Sequence 6, Appli
20	595	58.2	212	9	US-09-996-738-5	Sequence 5, Appli
21	594	58.1	214	12	US-10-625-260-6	Sequence 6, Appli
22	594	58.1	214	12	US-10-625-260-9	Sequence 9, Appli
23	594	58.1	214	13	US-10-061-658-6	Sequence 64, Appli
24	594	58.1	214	13	US-10-061-658-9	Sequence 5, Appli
25	594	58.1	214	16	US-10-474-832-64	Sequence 5, Appli
26	593	58.0	214	12	US-10-625-260-5	Sequence 5, Appli
27	593	58.0	214	13	US-10-061-658-5	Sequence 5, Appli
28	593	58.0	214	16	US-10-474-832-63	Sequence 531, App
29	590	57.7	1167	16	US-10-741-601-531	Sequence 532, App
30	590	57.7	1177	16	US-10-741-601-532	Sequence 5, Appli
31	585	57.2	195	10	US-09-805-354-5	Sequence 5, Appli
32	585	57.2	195	12	US-09-758-493-5	Sequence 5, Appli
33	585	57.2	195	14	US-10-144-259-5	Sequence 5, Appli
34	583	57.0	192	16	US-10-474-832-59	Sequence 59, Appli
35	580	56.8	192	16	US-10-474-832-61	Sequence 61, Appli
36	575	56.3	192	16	US-10-474-832-60	Sequence 60, Appli
37	574	56.2	195	10	US-09-805-354-7	Sequence 7, Appli
38	574	56.2	195	12	US-09-758-493-7	Sequence 7, Appli
39	574	56.2	195	14	US-10-144-259-7	Sequence 7, Appli
40	565	55.3	192	12	US-10-346-863-57	Sequence 57, Appli
41	478	46.8	103	9	US-09-764-870-472	Sequence 1193, Ap
42	478	46.8	103	11	US-09-764-875-1193	Sequence 472, App
43	478	46.8	103	14	US-10-125-540-472	Sequence 42, Appli
44	465	45.5	1147	12	US-10-336-603A-42	Sequence 187, App
45	465	45.5	1181	12	US-10-211-462-187	

ALIGNMENTS

RESULT 1

US-09-764-870-313
; Sequence 313, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ14
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-313

Query Match 100.0%; Score 1022; DB 9; Length 707;
Best Local Similarity 100.0%; Pred. No. 1.4e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CQYMDIVILVLDGNSIYPWVEQHFILNLKFFIYGPQIQGVVQYGEDVVHFEHLND	60
DB	175	CQYMDIVILVLDGNSIYPWVEQHFILNLKFFIYGPQIQGVVQYGEDVVHFEHLND	234
QY	61	YRSVKDVBAAASHIEORGGTGTATAGIEFARSEAFQKGRGAKKVMIVITDGESHDS	120
DB	235	YRSVKDVBAAASHIEORGGTGTATAGIEFARSEAFQKGRGAKKVMIVITDGESHDS	294
QY	121	DLEKVTQQSERNVTVYAVAVILGYNNRGINPETFLNEIKYIASDDDDKHHFNVITDEAAL	190
DB	295	DLEKVTQQSERNVTVYAVAVILGYNNRGINPETFLNEIKYIASDDDDKHHFNVITDEAAL	354
QY	181	KDIVDALGDRIFSLEGT	197
DB	355	KDIVDALGDRIFSLEGT	371


```
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/371,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: Curaseq version 0.1
; SEQ ID NO 6
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-6

Query Match      100.0%; Score 1022; DB 12; Length 1120;
Best Local Similarity 100.0%; Pred. No. 2.7e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGPGQIQGVVQYGEDVWVHEFLND 60
DB 90 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGPGQIQGVVQYGEDVWVHEFLND 149
QY 61 YRSVKDVVEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITGESHDSP 120
DB 150 YRSVKDVVEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITGESHDSP 209
QY 121 DLEKVIQSSRDNVTRYAVAVLGYNNRRGINPETFLEIKYIASDPDDKHFFNVVDEAAL 180
DB 210 DLEKVIQSSRDNVTRYAVAVLGYNNRRGINPETFLEIKYIASDPDDKHFFNVVDEAAL 269
QY 181 KDIVDALGDRIFSLEGT 197
DB 270 KDIVDALGDRIFSLEGT 286
```

```
RESULT 6
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338
```

```
Query Match      100.0%; Score 1022; DB 15; Length 1188;
Best Local Similarity 100.0%; Pred. No. 2.9e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGPGQIQGVVQYGEDVWVHEFLND 60
DB 159 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKFFYIGPGQIQGVVQYGEDVWVHEFLND 218
QY 61 YRSVKDVVEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITGESHDSP 120
DB 219 YRSVKDVVEAASHIEQGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITGESHDSP 278
QY 121 DLEKVIQSSRDNVTRYAVAVLGYNNRRGINPETFLEIKYIASDPDDKHFFNVVDEAAL 180
DB 279 DLEKVIQSSRDNVTRYAVAVLGYNNRRGINPETFLEIKYIASDPDDKHFFNVVDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
DB 339 KDIVDALGDRIFSLEGT 355
```

```
RESULT 7
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1188
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-810

Query Match      100.0%; Score 1022; DB 15; Length 1188;
Best Local Similarity 100.0%; Pred. No. 2.9e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKPFYIGPGQIQGVGVQYGEDVWVHEFLND 60
Db 159 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKPFYIGPGQIQGVGVQYGEDVWVHEFLND 218
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKWMIVITDGESHDP 120
Db 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKWMIVITDGESHDP 278
QY 121 DLEKVIQOSERDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 180
Db 279 DLEKVIQOSERDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
Db 339 KDIVDALGDRIFSLEGT 355

RESULT 8
US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF4892
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-35

Query Match      100.0%; Score 1022; DB 10; Length 1189;
Best Local Similarity 100.0%; Pred. No. 2.9e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKPFYIGPGQIQGVGVQYGEDVWVHEFLND 60
Db 159 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKPFYIGPGQIQGVGVQYGEDVWVHEFLND 218
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKWMIVITDGESHDP 120
Db 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKWMIVITDGESHDP 278
QY 121 DLEKVIQOSERDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 180
Db 279 DLEKVIQOSERDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
Db 339 KDIVDALGDRIFSLEGT 355

RESULT 9
US-09-836-353A-35
; Sequence 35, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF4891
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-35

Query Match      100.0%; Score 1022; DB 10; Length 1189;
Best Local Similarity 100.0%; Pred. No. 2.9e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKPFYIGPGQIQGVGVQYGEDVWVHEFLND 60
Db 159 CQYMDIVIVLDGNSIYPWVEVQHFLNLIKPFYIGPGQIQGVGVQYGEDVWVHEFLND 218
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKWMIVITDGESHDP 120
Db 219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKWMIVITDGESHDP 278
QY 121 DLEKVIQOSERDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 180
Db 279 DLEKVIQOSERDNVTRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVDEAAL 338
QY 181 KDIVDALGDRIFSLEGT 197
Db 339 KDIVDALGDRIFSLEGT 355

RESULT 10
US-10-262-839-4
; Sequence 4, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John.
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
```

```
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; REMAINING PRIOR APPLICATION DATA REMOVED - SEE FILE WRAPPER OR PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-262-839-4

Query Match      100.0%; Score 1022; DB 12; Length 1189;
Best Local Similarity 100.0%; Pred. No. 2.9e-98;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CQYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVFHFLND 60
Db      159 CQYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVFHFLND 218

Qy      61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 120
Db      219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 278

Qy      121 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180
Db      279 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 338

Qy      181 KDIVDALGDRIFSLEGT 197
Db      339 KDIVDALGDRIFSLEGT 355

RESULT 11
US-10-260A-3386
; Sequence 3386, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3386
; LENGTH: 437
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
; US-10-108-260A-3386

Query Match      99.6%; Score 1018; DB 15; Length 437;
Best Local Similarity 99.5%; Pred. No. 1.9e-98;
Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CQYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVFHFLND 60
Db      159 CQYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVFHFLND 218

Qy      61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 120
Db      219 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 278

Qy      121 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180
Db      279 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 338

Qy      181 KDIVDALGDRIFSLEGT 197
Db      339 KDIVDALGDRIFSLEGT 355

RESULT 12
US-09-805-354-8
; Sequence 8, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805,354
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-805-354-8

Query Match      98.0%; Score 1002; DB 10; Length 193;
Best Local Similarity 100.0%; Pred. No. 3e-97;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CQYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVFHFLND 60
Db      1 CQYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVVQYGEDVVFHFLND 60

Qy      61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 120
Db      61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQKGRKGAKKVMIVITDGESHDS 120

Qy      121 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180
Db      121 DLEKVIQSERDNVTRYAVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDEAAL 180

Qy      181 KDIVDALGDRIFS 193
Db      181 KDIVDALGDRIFS 193

RESULT 13
US-09-758-493-8
; Sequence 8, Application US/09758493
; Publication No. US20040086935A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND
; FILE OF INVENTION: USES THEREOF
; FILE REFERENCE: 00786-804001
; CURRENT APPLICATION NUMBER: US/09/758,493
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-758-493-8

Query Match
Best Local Similarity 100.0%; Score 1002; DB 12; Length 193;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CQTYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVGVQYGEDVWHEFHND 60
Db 1 CQTYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVGVQYGEDVWHEFHND 60
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKVMIVITDGESHDP 120
Db 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQGGKRGAKKVMIVITDGESHDP 120
QY 121 DLEKVIQOSERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVTDAAAL 180
Db 121 DLEKVIQOSERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVTDAAAL 180
QY 181 KDIVDALGDRIFS 193
Db 181 KDIVDALGDRIFS 193

RESULT 15
US-09-984-130-103
; Sequence 103, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-103

Query Match
Best Local Similarity 58.6%; Score 599; DB 10; Length 1151;
Matches 116; Conservative 33; Mismatches 48; Indels 2; Gaps 1;

QY 1 CQTYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGPGQIQGVGVQYGEDVWHEFHND 60
Db 139 CSTQLDIVIVLDGNSIYPWDSVTAFLNLLKRMIDIGPKQTQVGIQYGENVTHEFNLNK 198
QY 61 YRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAF--QKGGKRGAKKVMIVITDGESHDP 118
Db 199 YSSTEEVIVAAKKIVQGGQTMTALGTDITARKEAFTEARGARGVKKVMIVITDGESHDP 258
QY 119 SPDLKVIQOSERDNTVRYAVAVLGYNNRRGINPETFLNEIKYIASDPDDKHFNFVTDAAAL 178
Db 259 NHRLLKVIQCCDENIQRFSAIILGYSYNNRGNLSLSTKEKVEEIKSIASEPTEKHFNFVSDLE 318
QY 179 ALKDIVDALGDRIFSLEGT 197
Db 319 ALVTIVTLGERIFALEAT 337

Search completed: September 21, 2004, 13:29:04
Job time : 36.493 secs
```